



Reeta Ojala

## **Kaupungin työntekijöiden liikeratojen seuranta ja hyödyntäminen kiinteistöhuollon toiminnassa**

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi  
diplomi-insinöörin tutkintoa varten.

Espoossa 05.04.2019

Valvoja: Professori Kirsi Virrantaus

Ohjaaja: DI Mikko Viitala

---

**Tekijä** Reeta Ojala

---

**Työn nimi** Kaupungin työntekijöiden liikeratojen seuranta ja hyödyntäminen  
kiinteistöhuollon toiminnassa

---

**Maisteriohjelma** Geoinformatics

---

**Koodi** ENG22

---

**Työn valvoja** Professori Kirsi Virrantaus

---

**Työn ohjaaja** DI Mikko Viitala

---

**Päivämäärä** 05.04.2019

---

**Sivumäärä** 70+34

---

**Kieli** suomi

---

### Tiivistelmä

Työn tavoitteena oli hyödyntää spatio-temporaalista aineistoa eli kiinteistöhuollon työntekijöiden sijaintia ja työtehtäviä tiedonvälityksen parantamiseksi ja hyödyntää kerättyä aineistoa mahdollisimman paljon erilaisilla menetelmillä ja ArcGIS-ohjelmistoilla. Tapaustutkimus tehtiin Porin kaupungille ja tavoite oli automatisoida tekstiviesti-ilmoitukset huoltotoimenpiteen ilmoittajalle sekä tehdä tiedonlöytämisen prosessi spatio-temporaalisen aineiston tiedon louhintaan.

Porin kaupunki käyttää ArcGIS-ohjelmistoja aineiston keruuseen, analysoimiseen ja tallentamiseen. Aineistoa oli kerätty työntekijöiltä heidän antamalla suostumuksella jo ennen tämän työn aloittamista. Tapaustutkimuksessa käytetään Workforce for ArcGIS -sovelluksessa kertynyttä aineistoa hyödyksi mahdollisimman paljon.

Tapaustutkimuksessa käytettiin tiedon löytämisen prosessia aineiston louhintaan. Menetelmiksi valikoitui klusterointi, luokittelu ja visuaaliset menetelmät. Klusteroinnin algoritmeja olivat DBSCAN, HDBSCAN ja OPTICS. Luokittelun menetelmät olivat päätöspuu ja luokittelusäännöt. Visuaaliset menetelmät olivat aika-avaruuskuutiot liikerata-aineistona sekä pisteiden ryhmittäminen aikasarjakuutioihin viikko- ja tuntitarkastelulla.

Tapaustutkimuksessa tehtiin suunnitelma ja selvitys tekstiviestien lähettämisen automatisointia varten. Työssä käytettiin klusterointialgoritmeja suosittujen sijaintien löytämiseen. Liikerata-aineisto esikäsiteltiin ja luokiteltiin kolmeen luokkaan, kiinteistöön, tieverkkoon ja tuntemattomaan. Liikeradat sisältävät tiedon myös ajomatkasta ja työajasta. Lisäksi tehtiin aika-avaruuskuutioanalyysyjä, jotka osoittivat nousevia ja laskevia muutostrendejä huollettavissa kiinteistöissä sekä kuuma- ja kylmäpisteitä, jotka kuvastivat kiinteistöjen ja työntekijöiden työskentelytapoja. Työn avulla

Porin kaupunki pystyi parantamaan tiedonkulkua työntekijöiden välillä, tehostamaan työnjakoa ja analysoimaan tulevaisuuden remonttitarpeita kaupungin kiinteistöjen huoltotoimenpiteitä seuraten.

---

**Avainsanat** Paikkatieto, reaaliaikainen tieto, spatio-temporaalinen, liikeradat, tiedon löytämisen prosessi, työntekijän seuranta, kiinteistöhuolto, aika-avaruuskuutio, louhinta

---

<b>Author</b> Reeta Ojala		
<b>Title of thesis</b> Trajectory tracking and utilization for real estate maintenance workers		
<b>Master programme</b> Geoinformatics	<b>Code</b> ENG22	
<b>Thesis supervisor</b> Professor Kirsi Virrantaus		
<b>Thesis advisor</b> Mikko Viitala MSc		
<b>Date</b> 04.05.2019	<b>Number of pages</b> 70+34	<b>Language</b> Finnish

**Abstract**

The aim of this research was to discover variable data mining methods for spatial-temporal data that is worker location tracking data and assignments. The data was used to enhance communication and to discover information with knowledge discovery process and ArcGIS software. A case study is for the City of Pori and the first objective was to automate a notification system for real estate maintenance. The notification system's aim is to send a SMS message each time the status of the maintenance task is changed. The second objective was to do a knowledge discovery process in data.

The City of Pori utilizes ArcGIS-software for data collection, analysis and storage. The data has been collected from the real estate maintenance workers with their consent and it is used in this research. The case study utilized data from Workforce for ArcGIS -mobile application as much as possible.

The case study utilized the data by performing the knowledge discovery process in data. During the process the methods are chosen, and they are clustering, classification and visual methods. The algorithms used in clustering are DBSCAN, HDBSCAN and OPTICS. The classification methods include decision making tree and IF-THEN rules. The visual methods are space-time cube from trajectory data and from points into time series bins.

In the case study the first objective included descriptions and results of the investigation of automating notification system using test data. Work for the second objective included pre-processed and classified trajectory data into three classes: Real estate, Travel by car and Unknown. The trajectories had information of how much driving and work hours have been used in real estate maintenance for different time periods or areas. In addition, the study case included space-time cube analytics, that determine trending and emerging hot spot real estate, according to collected points in the area and point of time of the real estate.

The City of Pori can utilize these results by enhancing communication between workers and relevant people, optimize work load distribution and analysing the future needs for renovation of the real estate the city owns.

---

**Keywords** GIS, real-time data, spatial-temporal data, trajectory data, knowledge discovery process in data (KDD), worker tracking, real estate maintenance, space-time cube, data mining

---

## Alkusanat

*Haluan antaa kiitoksen Porin kaupungille ja erityisesti ohjaajalleni Mikko Viitalalle työn aiheesta ja mielenkiintoisesta projektista. Projekti sai alkunsa tarpeesta kehittää ja helpottaa ihmisen työtä ja sitä kautta laajentaa jo olemassa olevia aineistoja ja palveluita paremmin käyttöön.*

*Haluan kiittää työni valvojaa professori Kirsi Virrantausta, jonka ansiosta sain uuden kipinän ja innon työn tekemiseen ja rutkasti tukea ja apua matkan varrella.*

*Erityiskiitos myös Turkalle ja isälleni, jotka jaksoivat katsoa pitkää prosessia ja auttaa aina pyyteettömästi loppuun asti.*

Espoo 05.04.2019



Reeta Ojala



# Sisällysluettelo

Tiivistelmä	i
Abstract	ii
Alkusanat	iii
Sisällysluettelo	1
Lyhenteet	2
1 Johdanto	3
1.1 Työn rakenne	4
1.2 Tutkimusmenetelmät	4
2 Teoria ja aikaisemman tiedon kuvaus	5
2.1 Liikeratojen tiedon louhinta	5
2.1.1 Liikeratatiedon esikäsittely	8
2.1.2 Klusterointimenetelmät	9
2.1.3 Luokittelumenetelmät	11
2.1.4 Visuaaliset menetelmät	13
3 Tutkimusaineisto ja tutkimusmenetelmät	16
3.1 Tutkimusaineisto	16
3.2 Tutkimusmenetelmät	18
3.2.1 Tutkimuksessa käytettävät ohjelmistot	19
4 Tutkimustulokset	22
4.1 Tapaustutkimuksen ensimmäinen tutkimuskysymys	22
4.2 Tiedon löytämisen prosessi liikeratojen hyödyntämisessä	23
4.2.1 Klusterointimenetelmien tulokset	33
4.2.2 Luokittelumenetelmien tulokset	35
4.2.3 Visuaalisten menetelmien tulokset	46
5 Tulosten arviointi ja vertailu	60
5.1 Tulosten arviointi	60
6 Johtopäätökset	65
7 Yhteenveto	66
Lähdeluettelo	67
Liiteluettelo	70
Liitteet	

## Lyhenteet

DBSCAN	engl. Density Based Spatial Clustering of Applications with Noise
GNSS	Satelliittipaikannusjärjestelmä (engl. Global Navigation Satellite System)
GPS	Maaailmanlaajuinen paikallistamisjärjestelmä (engl. Global Positioning System)
GSM	maailmanlaajuinen matkapuhelinverkko (engl. Global System for Mobile Communication)
KDD	Tiedon louhintaprosessi (engl. Knowledge Discovery in Data)
netCDF	engl. Network Common Data Form
OPTICS	engl. Ordering Points to Identify the Clustering Structure
RFID	radiotaajuinen etätunnistus (engl. Radio Frequency Identification)
SaaS	engl. “Software as a Service”, tarkoittaa ohjelmiston hankkimista palveluna perinteisen lisenssipohjaisen tavan ja ohjelmiston asentamisen sijaan.
WLAN	langaton lähiverkkotekniikka (engl. Wireless Local Area Network)

# 1 Johdanto

Kaupungin kiinteistöhuoltotoiminta sisältää useita kiinteistöjä ja huoltotyöntekijöitä, jotka liikkuvat huollettavien kohteiden välillä. Huoltotoimenpiteitä on kahdenlaisia, on akuutteja tarpeita muutoksen tai rikkoutumisen takia ja on ennalta sovittuja rutiinitehtäviä, jotka ylläpitävät kiinteistöjä. Näiden tehtävien valvominen ja jakaminen huoltotyöntekijöille voi olla haastavaa, sillä huollettavat kohteet ja työntekijät toimivat koko kaupungin alueella kaukana toisistaan. Etäisyyden takia työnjohto voi hyödyntää paikkatietoa tehtävien jakamiseen ja seuraamiseen, sillä suurimmat vaikutukset toiminnassa on sijainti ja aika – missä, milloin ja kauan työtehtäviä suoritetaan.

Kiinteistöhuollontoiminnassa syntyy spatio-temporaalista aineistoa, joka on työntekijän sijainti kullakin ajanhetkellä sekä työtehtävien sijainti ja aika tehtävän aloittamisesta ja suorittamisesta. Kun paikkatietoa kerätään, sitä voidaan hyödyntää monella kiinteistöhuollon osa-alueella. Tämän aineiston keräämiseen pitää olla työntekijöiden suostumus. Suomen laki säätelee henkilöiden yksityisyyden suojasta työelämässä. Lain tarkoitus: *“Tämän lain tarkoituksena on toteuttaa yksityiselämän suojaa ja muita yksityisyyden suojaa turvaavia perusoikeuksia työelämässä”* (Finlex 13.8.2004/759). Laissa sanotaan, että työnantaja saa käsitellä vain välittömästi työntekijän työsuhteen kannalta tarpeellisia henkilötietoja. Yleisimmissä tapauksissa edellytyksenä on hankkia työntekijältä suostumus tietojen keräämiseen. (Finlex 13.8.2004/759.) Tässä työssä käytetään vain aineistoa, joka on kerätty työntekijän suostumuksella. Aineistoa käsitellään vain sovitulta työajalta kerättyä tietoa, jotta työntekijä voi kontrolloida keräystä itse, sillä aineistoin keruu tapahtuu älypuhelimella.

Työn tapaustutkimuksen kohteena on hyödyntää spatio-temporaalista aineistoa eli kiinteistöhuollon työntekijöiden sijaintia ja työtehtäviä tiedonvälityksen parantamiseksi ja hyödyntää kerättyä aineistoa mahdollisimman paljon erilaisilla menetelmillä ja ArcGIS-ohjelmistoilla.

Työssä käytetty aineiston keräys on aloitettu jo aiemmin ja noudattaa Esrin Workforce for ArcGIS -sovelluksen antamaa tietokantarakennetta. Tässä työssä käytetään reaaliaikaista aineistoa ja ennestään kerättyä historiallista aineistoa.

Porin kaupungin tavoite tapaustutkimuksessa on hyödyntää reaaliaikaista spatio-temporaalista aineistoa työtehtävien automatisointiin ja kustannuslaskentaan. Automatisointi sisältää tekstiviestien lähettämisen silloin, kun suoritettava työtehtävän tila muuttuu esimerkiksi silloin, kun tehtävä aloitetaan ja silloin, kun tehtävä on valmis. Kustannuslaskenta perustuu työntekijöiden käyttämään aikaan huollettavissa kohteissa, jotka ovat Porin kaupungin omistuksessa olevat kiinteistöt. Aika lasketaan työntekijän sijainnin perusteella keskimäärin minuutin tarkkuudella ja puhelimen satelliittipaikannusjärjestelmän (GNSS) tarkkuudella kiinteistörajajen mukaan. Lisäksi voidaan tarkkailla, paljonko työntekijän aikaa kuluu huoltokohteiden ulkopuolella ja kohteiden välillä kulkemiseen ja esimerkiksi laskea ajomatkojen pituudet.

Työn keskeisimpiä pohdintoja ovat, miten tietoa louhitaan, mitkä analyysityökalut ja tulokset hyödynnetään ja miten erilaiset analyysit tuottavat hyötyä Porin kaupungille. Rajoitteita tuottaa esimerkiksi Suomen laki, sillä aineisto sisältää työntekijöiden sijainnin seuranta. Lisäksi rajoituksena on käyttää vain ArcGIS-ohjelmistoa työn tekemiseen.

Työn tutkimuskysymykset ovat:

1. Miten voidaan automatisoida ilmoituksen lähettäminen työtehtävän tilan muuttuessa?
2. Miten Porin kaupungin kiinteistöhuollossa kerättävää aineistoa voidaan hyödyntää kiinteistöhuollon toiminnassa?

## **1.1 Työn rakenne**

Ensimmäinen luku on johdanto, joka esittelee työn taustatiedot, tapaustutkimuksen tavoitteet ja tutkimuskysymykset. Toisessa luvussa esitellään kirjallisuuskatsaus tiedon löytämisen prosessista ja sen menetelmistä, jotka ovat klusterointi, luokittelu ja visuaaliset menetelmät. Kolmannessa luvussa esitellään käytössä olevat aineistot, menetelmät ja ohjelmistot, joita käytetään tapaustutkimuksessa. Neljännessä luvussa esitellään tapaustutkimuksen tulokset. Viidennessä luvussa on tutkimukseen liittyviä pohdintoja ja tulosten arviointia. Kuudes luku on johtopäätökset ja seitsemäs luku on työn yhteenveto.

## **1.2 Tutkimusmenetelmät**

Työssä sovelletaan konstruktivistista tutkimusmenetelmää, jossa pyritään ratkaisemaan kiinteistöhuollon työmäärän seuranta kiinteistökohtaisesti ja aineiston louhinta mahdollisimman laadukkaasti sekä mittaamaan sen onnistumista. Tutkimusmetodina on kirjallisuuskatsaus, jonka pohjalta pyritään tekemään perustellut valinnat käytettävien metodien osalta. Empiirisessä tutkimuksessa toteutetaan tapaustutkimus, joka sisältää työtehtävien automatisointia aineiston avulla sekä aineiston louhintaa kirjallisuuskatsauksessa löytyneillä algoritmeilla ja tutkitaan niiden hyötyä esittelemällä ne aineiston omistajalle, Porin kaupungille. Tapaustutkimuksessa käytetään ArcGIS-ohjelmistoa.

## 2 Teoria ja aikaisemman tiedon kuvaus

Tällä vuosikymmenellä kerätään valtava määrä tietoa, niin ihmisistä, kulkuvälineistä, rakennuksista kuin laitteistakin. Asioiden kehitystä voidaan seurata reaaliaikaisesti, mutta tiedon määrä on niin suuri, ettei sitä voida hyödyntää sellaisenaan. Liikerata on yksi aineistomuoto, jota voidaan hyödyntää spatio-temporaalisen aineiston louhinnassa. Tässä luvussa esitellään liikeratojen tiedon louhintamenetelmät.

### 2.1 Liikeratojen tiedon louhintamenetelmät

Liikerata on aikaan ja paikkaan sidottua aineistoa, joka voidaan kerätä käyttämällä paikannusmenetelmänä GNSS:ää, GSM:ää, RFID:tä tai WLAN:a. Paikannusmenetelmä vaikuttaa huomattavasti aineiston tarkkuuteen ja luotettavuuteen, mikä tulee ottaa huomioon jo aineiston keräyksessä. Liikerata voi kuvata esimerkiksi auton, ihmisen, eläimen tai myrskyn etenemistä ajan ja paikan suhteen. Liikerata voidaan yleisesti esittää kaavalla:

$$T = \langle p_1 \dots p_n \rangle$$

missä  $p_k = (id_k, loc_k, t_k, A_k)$  on järjestyksessä kohdassa  $k$ ,

missä  $id_k$  on sijainnin tunnistenumero,

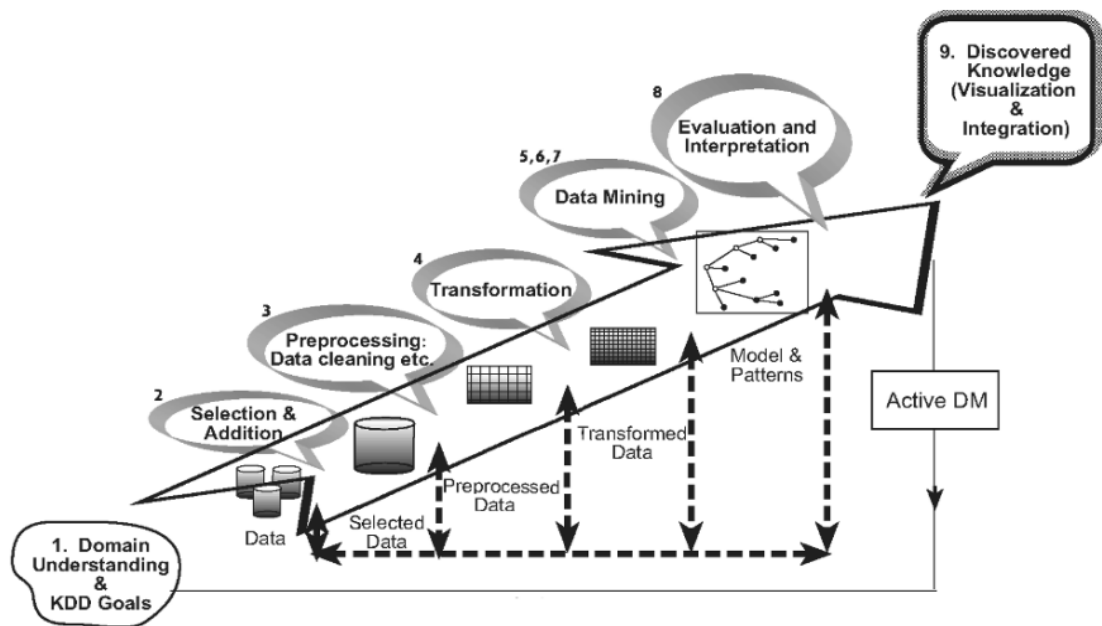
$loc_k$  on sijaintitieto (esimerkiksi XY tai solutunniste)

$t_k$  on ajanhetki sijainnille

$A_k$  on lista muista ominaisuustiedoista kuten suunta, nopeus tai tila  
(Mazimpaka et al. 2016, s. 63).

Liikeratojen louhinnan tarkoitus on tuoda esiin tietoa ilmiöistä tai kohteista. Tietoja voidaan esittää esimerkiksi numeerisesti tai visuaalisesti (Bédard & Han, 2009). Liikeratoja louhitaan ensisijaisesti klusteroimalla ja luokittelemalla. Lisäksi voidaan käyttää visuaalisia menetelmiä. Louhinnalla voi löytyä poikkeavia sijainteja ja ilmiöitä, joita voidaan hyödyntää käyttäytymistutkimuksissa. (Mazimpaka et al. 2016. s. 64.)

Aineiston louhinta on tärkeä osa tiedon löytämisen prosessia (Knowledge Discovery in Data, KDD), missä erotellaan aineistosta käytettävä tieto. Tiedon löytämisen prosessissa (KDD) on Maimon et al. (2010) mukaan yhdeksän vaihdetta. Vaiheet on kuvattu seuraavassa kuvassa (Kuva 1).



Kuva 1: Tiedon löytämisen prosessi (Maimon et al. 2010. s.3).

Kuvassa (Kuva 1) näkyy, miten tiedon löytämisen prosessi on jatkuvaa uudelleen arviointia ja paluuta takaisin aiempiin vaiheisiin. Prosessissa on kuitenkin selkeästi yhdeksän vaihetta, jotka ovat kuvattu seuraavalla tavalla.

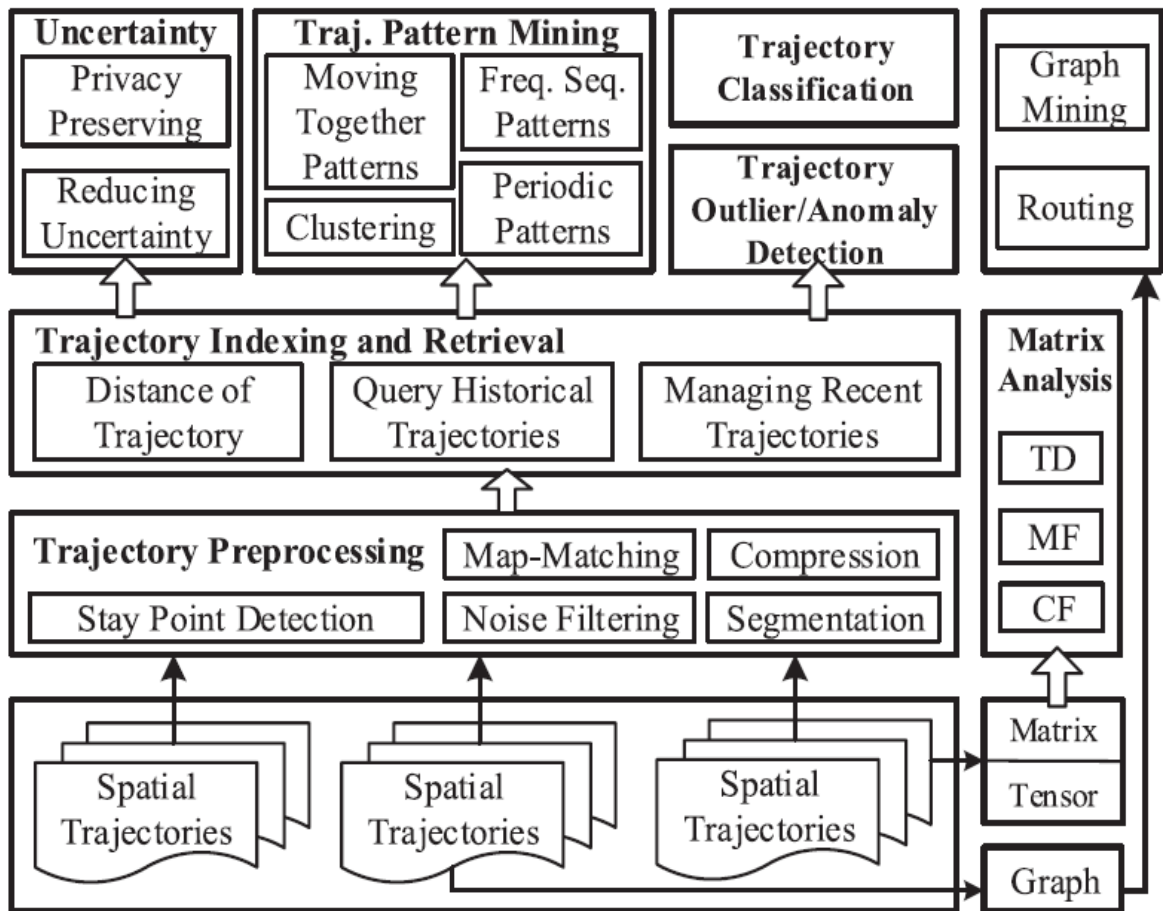
Tiedon löytämisen prosessin vaiheet (Maimon et al. 2010) :

1. Määritellään tavoitteet ja työskentely-ympäristö
2. Kerätään ja valitaan aineistoja. Tässä vaiheessa kuuluu selvittää, mitä aineistoja on valmiina, mitä on mahdollista saada lisää. Lisäksi tarkistetaan, että on riittävän paljon erilaisia ominaisuustietoja, jotta prosessi ei kaadu myöhemmissä vaiheissa (Miller & Han. 2009, s.6). Ominaisuuksien lisäksi aineiston sisäinen tarkkuus tarkistetaan viiden kriteerin mukaisesti (1) sijaintitarkkuus, (2) ominaisuustietotarkkuus, (3) temporaalinen tarkkuus, (4) looginen eheys ja (5) valmius ja täydellisyys (Veregin. 1998, Gupta ja Morrison 1995).
3. Käsitellään aineistoa laadun ja tarkkuuden parantamiseksi. Aineistokäsittely on kohinan ja poikkeavien havaintojen poistaminen ja mahdollisten ominaisuustietojen täydentäminen. Tämä vaihe saattaa muuttua prosessin aikana kaikista aikaa vievimmäksi vaiheeksi. Lisäksi aineiston käsittelijän tulee tiedostaa aineiston ominaisuudet ja rajoitukset, sillä muuten aineiston väärinkäyttö tai väärät tulkinnot voivat aiheuttaa vakavia laillisia, sosiaalisia tai ekonomisia vaikutuksia aineiston hyödyntäjälle (Devillers et al. 2002).
4. Muunnetaan aineisto. Aineiston muunnos on hyvin projektikohtainen ja voi tarkoittaa aineiston rajaamista, näytteiden rajaamista tai ominaisuuksien muunnosta esimerkiksi luokittelua hyödyntäen. Muunnoksen hyödyllisyys ja tarpeellisuus selkeytyy ajan myötä, kun tiedon löytämisen prosessia käydään läpi, sillä mitä enemmän aineistoa tutkii, sitä enemmän muunnostarpeita löytää.
5. Valitaan aineiston louhintamenetelmät. Aineiston louhintaan voidaan käyttää esimerkiksi luokittelua tai klusterointia. Tiedon louhinnan tavoitteena on ennustaa tai kuvailla aineistoa.

6. Valitaan algoritmit. Algoritmin valintaan vaikuttaa tiedon löytämisen tavoitteet; tarvitaanko tarkkaa tietoa vai enemmän ymmärrystä. Algoritmivalintoja erityisesti liikeratojen louhimisessa käsitellään tulevissa luvuissa.
7. Ajetaan algoritmit. Analyysin suorittamiseen saattaa mennä useampi ajo, sillä parametrien asetuksilla voidaan optimoida tuloksia.
8. Arvioidaan tuloksia. Tuloksia arvioidaan ja tulkitaan tiedon löytämisen tavoitteiden valossa. Tuloksista tulkitaan niiden hyödyllisyys ja ymmärrettävyys ja samalla dokumentoidaan tulokset.
9. Hyödynnetään löydettyä tietoa ja ymmärrystä. Hyödyntäminen voi tarkoittaa tuloksien visualisointia ja integrointia dynaamiseen aineistoon. Viimeisessä vaiheessa saattaa olla haasteista, sillä löydetty tieto ja mallit sovelletaan käytäntöön, eikä vain staattiseen otokseen aineistosta.

Jotta tiedon löytämisen prosessi voidaan saattaa loppuun, pitää olla ymmärrys erilaisista louhintamenetelmistä, jotka valitaan vaiheissa viisi. Menetelmät hyödyntävät erilaisia algoritmeja. Nämä algoritmit valitaan ja ajetaan vaiheissa kuusi ja seitsemän.

Kuvassa (Kuva 2) näkyy liikeratojen louhimisen paradigma. Kuvassa näkyvät vaiheet käsitellään luvuissa 2.1.1-2.1.3. Zhengin (2015, s. 3) mukaan liikeratatyyppejä on kaksi: historiallinen ja viimeaikainen, jotka tarvitsevat erilaiset käsittelymetodit.



Kuva 2: Paradigma liikeratojen aineiston louhimisesta (Zheng, 2015, s.2)

### 2.1.1 Liikeratatiedon esikäsittely

Liikeratoja voidaan louhia, kun aineisto on esikäsitelty. Esikäsitellyllä tarkoitetaan aineiston siivoamista, liikkumattomien ajanhetkien löytämistä (engl. stay point detection), liikeratojen kompressiota, karttasovitusta ja liikeratojen segmentointia. Aineiston siivoamisen tavoite on poistaa kohinaa, jotka johtuvat signaalin heikkoudesta aineiston keräyshetkellä, esimerkiksi rakennuksen sisällä. Liikkumattomien ajanhetkien löytämisessä käytettävä algoritmi tunnistaa sijainnit, missä liikkuva kohde on pysynyt paikallaan hetken tiettyjen raja-arvojen puitteissa, esimerkiksi, kun seurattava käy ravintolassa syömässä. Liikeratojen segmentointi jakaa liikeradan merkityksellisiin osiin aikajakson, spatiaalisen muodon tai semantiikan mukaan jatkojalostusta varten. Klusteroinnilla ja luokittelulla voidaan myös tehdä aineiston esikäsittelyä, esimerkiksi kohinan ja häiriöiden siivoamista aineistosta. (Zheng, 2015. s.2.)

Zheng et al. (2010, s.8) käyttivät liikeratojen segmentoinnissa ”change point detection segmentation” menetelmää, eli muutospisteensegmentointia. He tarkastelivat neljää eri kulkumuotoa: kävelen, autolla, linja-autolla ja pyörällä kulkemista. Muutospisteensegmentointi tarkoittaa, että liikeradasta löydetään segmentit, jotka ovat todennäköisemmin kulkumuodon muutosta, esimerkiksi autolla ajamisen muuttamista kävelemiseen. Zheng et al. (2010) päättelivät seuraavat kaksi tapausta heidän aineistonsa ja kokemuksensa pohjalta:

- Käveleminen on muutosvaihe kaikkien kulkumuotojen välissä. Toisin sanoen aloituspiste ja lopetuspiste kävelysegmentillä ovat suurella todennäköisyydellä muutospisteitä.
- Tyypillisesti ihmisten pitää ensin pysähtyä ja sitten liikkua uudelleen, kun he muuttavat kulkumuotoaan. Esimerkiksi osalla keräyspisteillä nopeus pitäisi olla lähellä nollaa näillä muutoshetkillä.

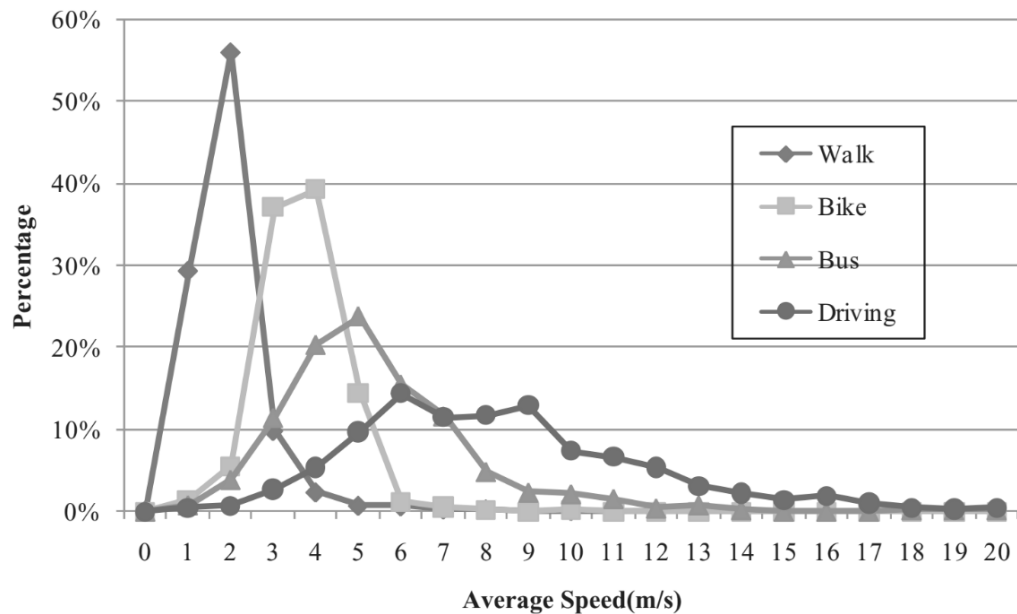
Näin ollen, kävelysegmenttisen erottelu ja tunnistaminen aineistosta pilkkoo aineiston segmentteihin, jotka kuvastavat suurella todennäköisyydellä eri kulkumuotoja.

*Taulukko 1: Zheng et al. (2010, s.8) Muutostaulukko, joka esittää todennäköisyyden mihin kulkumuotoon ihminen vaihtaa.*

Transportation Modes	Walk	Driving	Bus	Bike
Walk	/	41.0%	49.0%	9.0%
Driving	99.7%	/	0%	0.3%
Bus	98.7%	0.8%	/	0.5%
Bike	99.8%	0%	0.2%	/

Taulukko 1 esittää muutosmatriisissa neljän kulkumuodon muutoksesta toiseen kulkumuotoon. Taulukko on yhteenveto aineistosta, joka on kerätty 65 henkilön voimin 10 kuukauden ajan. Melkein jokaisella kerralla ajamalla, linja-autolla ja pyörällä kulkumuodon muutos oli kävelyyn ennen kuin muutos oli johonkin toiseen. Vain erittäin harvoissa tapauksissa osa ihmisistä otti taksikyydin (autoilua) heti linja-autolla kulkemisen jälkeen. Seurattavat ihmiset merkitsivät itse omat kulkumuotonsa kerättävään GPS-aineistoon ja havaittiin, että on erittäin todennäköistä jättää merkitsemättä erittäin lyhyet kävelyvälit auto- ja linja-autokulkumuotojen välistä. Tästä syystä linja-autosta vaihtaminen autoiluun on tapahtunut 0.8% kerroista. Toisin sanoen, kävelysegmentti on todellisuudessa siinäkin välissä, vaikka onkin erittäin lyhyt etäisyydeltään. (Zheng et al. 2010.)





Kuva 3: Kerätyn aineiston keskinopeus eri kulkumuodoilla (Zheng et al. 2010, s.9)

Kuvassa (Kuva 3) näkyy kerätyn aineiston keskinopeudet ja niiden prosentuaalinen osuus koko sen kulkumuodon nopeuksista. Aineistosta nousee selkeästi neljän eri kulkumuodon nopeudet, mutta linja-autolla ja autolla ajaminen ovat jonkin verran samankaltaisia. Suurin osa kävelystä (noin 85 %) on alle 2 m/s ja enää 10 % on noin 3 m/s. Vaihtoehtoisesti autoilla ajamisessa on keskinopeudet jakautuneet tasaisemmin, noin 10 % jokaista nopeutta kohden 5 – 11 m/s välillä, eli yhteensä 70 % keskinopeuksista on 5–11 m/s (18–40 km/h). Kävelyn ja autoilun päällekkäisyys on vain parin prosentin verran aineistosta.

## 2.1.2 Klusterointimenetelmät

Tiedon löytämisen prosessin alalla klusterointianalyysit tunnetaan ohjaamattomana oppimisprosessina, sillä analysoitavasta aineistosta ei ole ennestään tietoa. (Birant ja Kut. 2006, s.208) Klusteroinnin tavoite on ryhmitellä aineisto liikekäyttäytymisen perusteella liikeratoihin. Ne, jotka ovat samassa klusterissa edustavat samanlaista liikekäyttäytymistä ja erilaista liikerataa kuin toisissa klustereissa. Liikeradoille on yritetty kehittää niille sopivaa klusterointimenetelmää, joka perustuu tilastotieteen ja todennäköisyyden malleihin. (Gaffney ja Smyth. 1999. s. 2.)

Klusterointialgoritmit voidaan luokitella viiteen pääryhmään, jotka ovat osittaiset, hierarkkiset, ruudukkopohjaiset, mallipohjaiset ja tiheyspohjaiset klusterointialgoritmit. (Birant ja Kut. 2006, s.209). Tiheyspohjaiset klusterointialgoritmit perustuvat ideaan, missä klusteri on alue missä kohteet ovat sijoittuneet tiheästi lähelle toisiaan. Nämä algoritmit käyttävät kiinteää etäisyyden mittaa, millä etsitään tiheitä alueita. Algoritmit etsivät tiheitä alueita, jotka erottuvat toisistaan harvojen alueiden välistä. Tiheyspohjaisilla klusterointialgoritmeilla löytää vain pyöreän muotoisia klustereita, sillä klustereiden etsimiseen käytetään kiinteää etäisyyttä (Han et al. 2009).

Yksi tiheyspohjainen klusterointialgoritmi on esimerkiksi DBSCAN, joka on kohinaa sisältävien kohteiden tiheyteen perustuva spatiaalinen klusterointialgoritmi. Tiheys

saadaan laskemalla pisteiden määrä yhden pisteen ympärillä tietyllä etäisyydellä. Pisteet, joiden tiheys on tietyn raja-arvon yli, tunnistetaan klustereiksi. DBSCAN on algoritmi, joka toimii hyvin isoilla aineistoilla ja se tunnistaa klustereita satunnaisella muodoilla eikä vaadi ennalta määriteltyä klustereiden lukumäärää. Kirjallisuudessa selviää, että DBSCAN on pitkään laajasti käytetty algoritmi erilaisissa tutkimuksissa. (Birant ja Kut. 2007. s. 209, Han et al. 2009. s.170.)

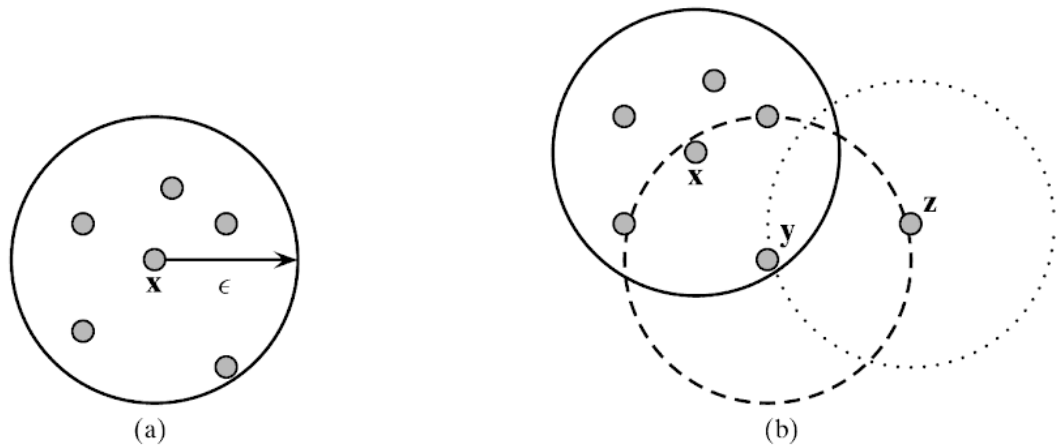


Figure 15.2. (a) Neighborhood of a point. (b) Core, border, and noise points.

Kuva 4: DBSCAN toiminnallisuus (Zaki ja Meira 2014.s. 376).

- a)  $\epsilon$  on naapurusalueen säde ja  $x$  on piste, missä tarkastelu tehdään.
- b) erilaiset pistetyypit, missä tiheyden raja-arvo on 6,  $x$  on keskipiste, sillä alueessa on 6 naapuripistettä,  $y$  on rajapiste, sillä sen ympärillä on alle 6 pistettä, mutta se on yksi  $x$ -pisteen naapureista, ja  $z$  on häiriöpiste.

DBSCAN on yksi nopeimmista klusterointimenetelmistä, mutta toimii parhaiten, jos on tarjolla selkeä hakuetaisyys, joka toimii hyvin potentiaalisten klustereiden kanssa. Tämä tarkoittaa, että kaikki merkittävät klusterit pitäisi löytyä samalla hakuetaisyydellä. (Ester et al. 1996.) Kuva 4 esittää tiheysalgoritmin toimintaperiaatetta, miten klusterin valinta toimii tietyllä hakuetaisyydellä ja tiheyden raja-arvolla. Kuvassa raja-arvo on 6, eli pisteitä pitää olla hakuetaisyyden sisällä 6, jotta se lasketaan klusteriksi (Zaki ja Meira 2014. s. 376).

DBSCAN algoritmista on toinen vaihtoehto, joka on itseään säätävä HDBSCAN. HDBSCAN käyttää joukkoa eri pituuksia erottelemaan eri tiheyksillä olevat klusterit kohinasta. HDBSCAN on tietolähtöisin klusterointimenetelmä ja täten vaatii vähiten käyttäjän antamia parametreja. (Camello et al. 2013).

Toinen suosittu tiheyspohjainen algoritmi on OPTICS, pisteiden järjestäminen klusterirakenteen tunnistamiseen, jonka käsitteet perustuvat DBSCAN algoritmiin ja siten tunnistaa klusterit. (Birrant ja Kut. 2007. s. 209.) OPTICS algoritmi käyttää vierekkäin olevien pisteiden etäisyyttä ja luo saavutettavuuskaavion, mitä hyödynnetään klustereiden erottamiseen kohinasta. OPTICS tarjoaa eniten joustavuutta klustereiden hienosäätämässä, mutta on laskennallisesti raskas erityisesti, kun on iso hakuetaisyys. OPTICS tuo ratkaisun DBSCAN algoritmin ongelmaa, missä DBSCAN epäonnistuu löytämään optimaalisia klustereita, kun aineisto on sekä tiheästi että harvasti jakautunut tarkasteltavalle alueelle (Han et al. 2009, s. 172). (Ankerst et al. 1999.)

### 2.1.3 Luokittelumenetelmät

Luokittelun tavoite on löytää säännöt, miten jakaa kohteet ennalta määrättyihin luokkiin. Toisin sanoen luokittelu on prosessi, missä etsitään malli tai funktio, joka kuvaa ja erottelee aineistoluokat (Han et al. 2012, s.18). Luokittelun perustana toimii ennalta määritetty luokkajako ja näyteaineisto, missä luokittelu on jo suoritettu. Tätä näyteaineistoa kutsutaan opetusjoukoksi. Luokittelu perustuu kohteiden ominaisuuksiin ja ennalta määritettyjen luokan ominaisuuksien yhdistämiseen. (Mazimpaka et al. 2016. s. 67.)

Suurin osa menetelmistä ovat kaksijakoisia. Ensin erotellaan osajoukko kohteista, jolla koulutetaan olemassa oleva luokittelumalli. Ensimmäiseen vaiheeseen kuuluu myös olennaisten ominaisuuksien tunnistaminen, millä luokittelu tehdään. Esimerkiksi ominaisuudet voivat olla keskinopeus, keskikiikhtyvyys, liikeradan tai tapahtuman kesto ja liikeradan pituus. Toisessa vaiheessa luokittelua valitaan standardiluokittelumalli valitulle aineistolle. (Mazimpaka et al. 2016. s. 67.)

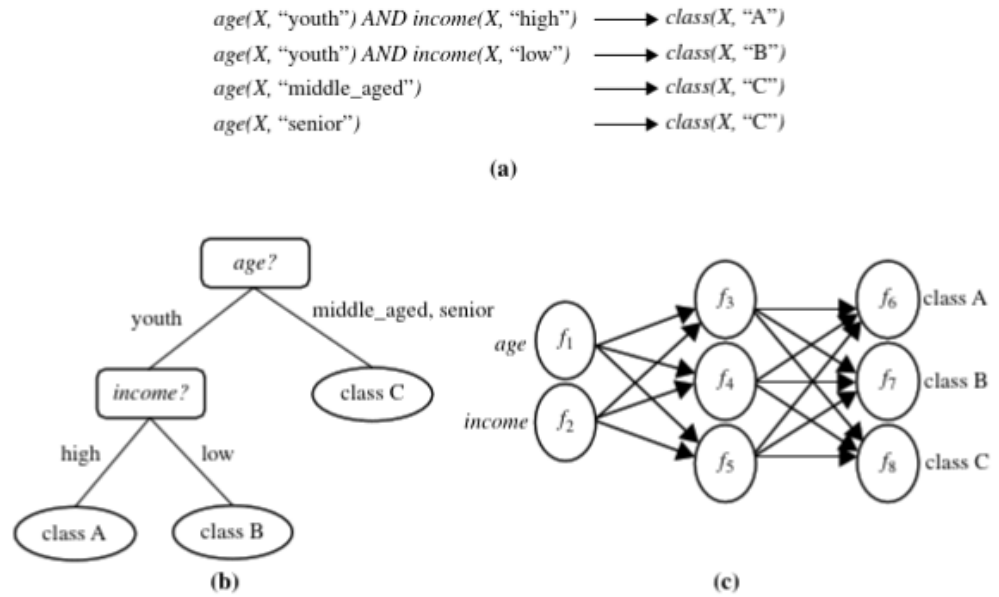
Nämä standardiluokittelumallit ovat esimerkiksi luokittelusäännöt (esim. IF-THEN säännöt), päätöspuumenetelmät ja neuroverkot. Muitakin malleja ovat esimerkiksi matemaattiset kaavat, Bayesläinen luokittelu, tukivektorikoneet ja k:n lähimmän naapurin luokittelumenetelmä. Osa näistä menetelmistä on esitetty kuvassa (Kuva 5), missä (a) on luokittelusäännöt, (b) on päätöspuumenetelmä ja (c) on neuroverkkomenetelmä. (Han et al. 2012, s.18-19.)

Yksinkertaisin luokittelumalli on luokittelusääntöjen käyttäminen. Siinä annetaan aineistojoukolle ehtolauseita ominaisuustietojen perusteella ja ehtojen mukaiset kohteet kuuluvat samaan luokkaan. Esimerkiksi kaikki kohteet, jotka ovat 10 metrin säteellä tieverkosta kuuluvat luokkaan: Tiet. Vaihtoehtoisesti kaikki kohteet, joiden nopeus on alle 10 km/h kuuluvat luokkaan: Jalankulkija. Kuten kaavassa (Kuva 5 (a)) esitetään, luokittelusäännöt ovat ominaisuustietojen perusteella olevia ehtoja kuten, ikäryhmä ja tuloluokka. Ehtoja voi olla yksi, kuten kaavassa on luokka C ”keski-ikäiset”, mutta voi olla myös useampia ehtoja, kuten kaavan luokka A ”Nuoriso” ja ”Korkea tuloiset”. (Han et al. 2012, s.18.)

Päätöspuumalli on astetta monimutkaisempi, sillä se on kulkukaavion omainen rakenne, missä jokainen solmukohta osoittaa testin aineiston attribuuttiarvoille, ja jokainen oksa edustaa testin tulosta. Puun lehdet edustavat luokkia. Päätöspuun avulla voidaan myös määrittellä luokittelusääntöjä. Kaavassa (Kuva 5 (b)) päätöspuu käsittelee useampaa testiä, kuten ikäryhmää ja tuloluokkaa. Haarojen pituudet voivat vaihdella, sillä tietylle luokalle voi olla vain yksi testi ja toiselle useampi, samalla tavalla kuin luokittelusäännöissä voi olla yksi tai useampi ehto. (Han et al. 2012, s.18-19.)

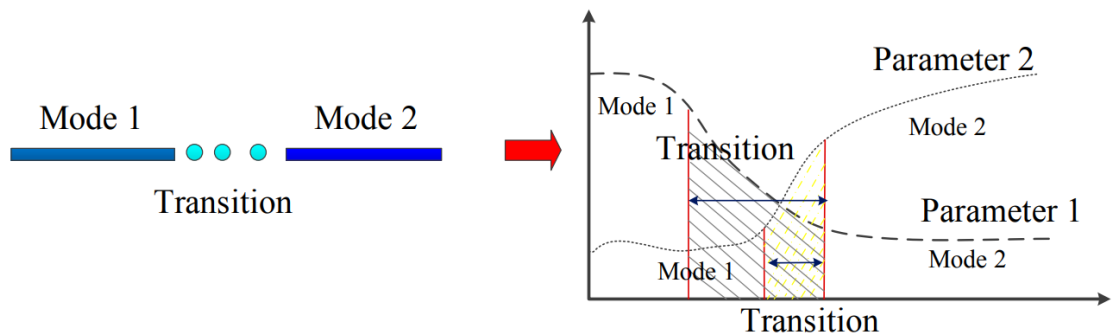
Neuroverkko luokittelumenetelmänä on joukko neuronimaisia prosessointiyksiköitä, joilla on painotettu yhteys yksiköiden välillä. Kaavassa (Kuva 5 (c)) neuroverkon alkupäässä on ikäryhmä ja tuloluokka, jolla tarkastellaan aineiston ominaisuuksia. Aineisto ajetaan neuroverkon läpi ja prosessointiyksiköt ohjaavat aineiston sille kuuluvaan luokkaan. (Han et al. 2012, s.19.)

Luokittelumenetelmä, mitä on käytetty liikeratojen luokitteluun, on esimerkiksi päätöspuu. Esimerkiksi Zheng et al. (2010, s. 5) luokitteli liikeratoja erilaisiin kulkuvälineisiin kuten kävelyyn, autoon, pyörään tai linja-autoon.



Kuva 5: Luokittelumallit eri muodoissa, (a) luokittelusäännöt eli IF-THEN malli, (b) päätöspuu, (c) neuroverkko. (Han et al. 2012, s.18.)

Liikeratasegmenttejä luokitellaan yleisesti vain historia-aineistosta kulkumuoto ja liikekäyttäytymisen tutkimuksissa (kulkumuotopäiväkirja-analyysi tai ihmisen liikekäyttäytymisen tiedon louhinta), sillä segmenttiaineistoa on vaikea luokitella reaaliaikaisesti. Reaaliaikaiset luokittelumenetelmät sovelletaan pisteaineistolle, jolloin voidaan luokittelun jälkeen muodostaa liikeratoja. Liikeratasegmenttien luokittelun haaste on muutospisteiden löytäminen, jotta kulkumuodon muutos löytyy esimerkiksi kävelystä autoiluun. Muutospisteet eivät todellisuudessa ole tarkkoja pisteitä, missä kulkijan nopeus muuttuu yhtäkkiä, vaan muutos on asteittainen, jossa hidastetaan tai kiihdytetään uuteen nopeuteen ja liikekäyttäytymiseen, joka on esitetty kaaviossa (Kuva 6). Tutkimuksissa on käytetty lähtöaineiston lisäksi lisätietoja, esimerkiksi parkkipaikkoja tai bussipysäkkejä, jotta luokittelun tarkkuus parane. Lisätietojen käyttäminen tuo lisävarmuutta luokitteluun, mutta myös vaatii enemmän laskentatehoa analyysia varten, joka tuo haasteita erityisesti reaaliaikaiseen luokitteluun. (Yang et al. 2018. s 9.)



Kuva 6: Muutospisteen löytäminen. Muutospiste on asteittainen. (Yang et. al 2018.)

### 2.1.4 Visuaaliset menetelmät

Visuaaliset menetelmät voivat olla aineiston kokonaiskuvan visualisointia, muutoksen visualisointia tai käyttäytymisen seurantaa ja tarkastelua. Visualisointimenetelmät voivat olla dynaamisia karttoja ja sovelluksia tai staattisia esityksiä aineistosta, jotka tuovat esiin haluttua informaatiota loppukäyttäjälle.

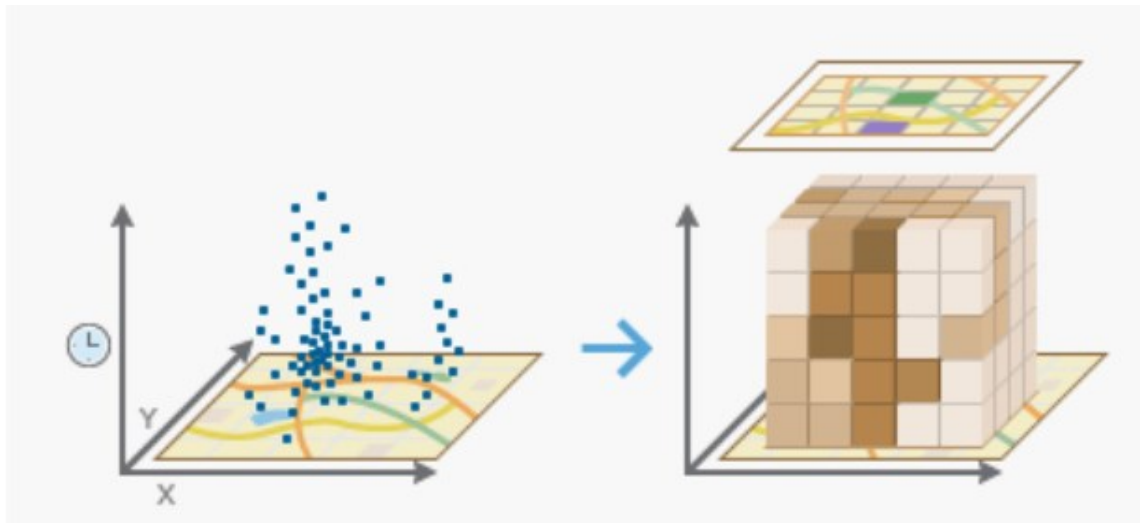
Spatio-temporaalista aineistoa voidaan visualisoida aika-avaruuskuutiolla, jota ensimmäisenä on ehdottanut visualisointimenetelmäksi Hägerstrand vuonna 1970. Hägerstrand ehdotti, että perinteisen kaksiulotteisen esitystavan lisäksi käytetään kolmatta ulottuvuutta esittämään ajan kulkua. Kokonaisuus on kolmiulotteisessa diagrammissa, joka näyttää miten ihmiset käyttäytyvät tilassa ja ajassa. Kuution pohja esittää maantieteellistä sijaintia ja kuution korkeus temporaalista ulottuvuutta. (Hägerstrand 1970.)

Aika-avaruuskuutio on hyödyllinen visualisointikeino, sillä se näyttää aineiston samanaikaisesti ja katsoja voi nähdä miten kohde muuttuu tai liikkuu ajan suhteen hävittämättä edellistä näkymää, mitä esimerkiksi aika-animaatioissa tapahtuu. Aika-animaatiot ovat monelle aineiston tulkitsijalle vaikeita, sillä niissä on haastavaa huomata eroja pidemmällä aikavälillä.

Tobler (1970, s. 236) muotoili maantieteen ensimmäisen lain seuraavalla tavalla: “...everything is related to everything else, but near things are more related than distant things.” Eli kaikki liittyy kaikkeen muuhun, mutta lähellä toisiaan olevat havainnot ovat suuremmalla todennäköisyydellä samankaltaisempia keskenään kuin kauempana toisistaan olevat havainnot. Esri Inc:n järjestämässä ja taltioimassa työpajassa L. Bennett et al. (2017) viittaavat tähän ensimmäiseen lakiin. He olivat tavanneet Toblerin ja kysyneet, mitä hän muuttaisi kirjoittamassaan väitteessään hänen nykytietämyksellään paikkatiedon analysoimisesta. Tobler oli vastannut, että lisäsi yhden muuttujan, ajan, mukaan väitteeseen seuraavalla tavalla: “... everything is related to everything else, but near *and recent* things are more related than distant things.” Eli lähellä toisiaan sijainnillisesti, että ajallisesti olevat havainnot ovat suuremmalla todennäköisyydellä samankaltaisempia keskenään kuin kauempana (sijainnin ja ajan suhteen) olevat kohteet. Toisin sanoen, on tärkeää ottaa huomioon paikkatietoaineiston arvot ja analyysit myös ajan suhteen. Aika-avaruuskuutio tuo tähän visualisointikeinon.

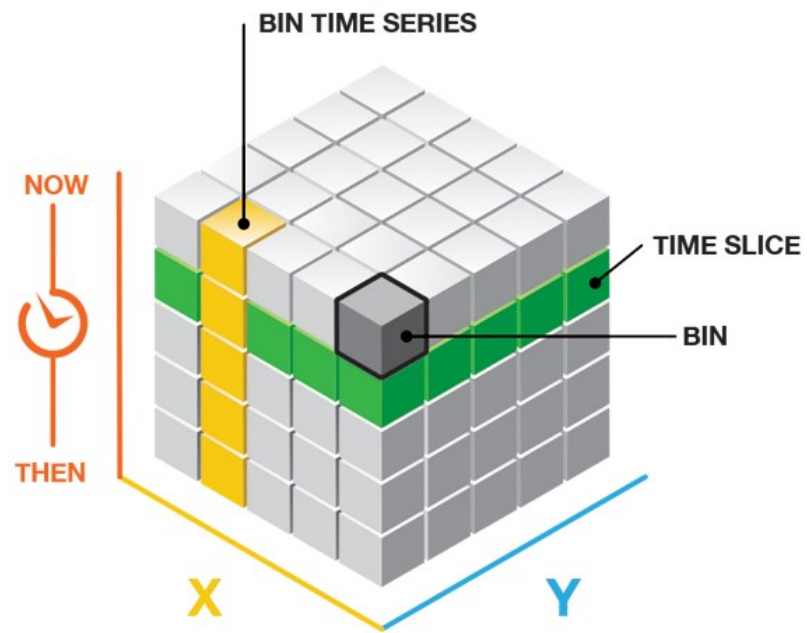
Esri Inc:n teknologialla aika-avaruuskuutiot luodaan kokoamalla pisteitä netCDF formaattiin, missä summataan ja tallennetaan aineisto moniulotteisessa rakenteessa. Käytettävä aineisto pitää olla piste, jolla on sijainti ja aikaleima. Aikaleiman avulla kuutioiden ajanjakso voidaan määritellä sekunniksi, minuutiksi, tunnit, päiväksi, viikoksi, kuukaudeksi tai vuodeksi. Ajanjaksoja pitää olla vähintään 10, eli päiväkohtaisessa jaossa aineistoa pitää olla 10 päivän ajalta. Kuva 7 esittää miten lähtöaineisto muokataan aika-avaruuskuutioksi. Aikavälin valitseminen ja aikavälin aloittamisen kohta on tärkeitä ottaa huomioon, sillä kuutiot aiheuttavat temporaalista vääristymää ajanjakson alussa ja lopussa. On tärkeää valita, lasketaanko kuutioiden ajanjakson alku aineiston ensimmäisestä kohteesta vai viimeisestä. Yleisin tapa on valita, että ajanjakso lasketaan uudemmasta aineistosta taaksepäin alkuun, sillä uudempi aineisto on usein myös tärkeämpi ja relevantimpi aineisto kun vanhempi (Miller and Han 2009, Roddick & Lees. 2009). Tämä pitää ottaa huomioon aika-avaruuskuutioiden tulkinnessa, sillä reunakuutioissa voi olla vähemmän aikaa kerättyä aineistoa. Esimerkiksi, jos aineisto ryhmitellään kuukauden välein ja aineistoa on kerätty 5.3.2016 alkaen ja päättyen

31.1.2019, niin maaliskuu 2016 kuukauden laatikko on jäänyt 4 päivää vajaaksi. (Esri, 2019.)



*Kuva 7: Kuvassa näkyy lähtöaineistona pisteitä, jotka on kuvattu XY-tasolla sijainnin suhteen ja Z-akselilla ajan suhteen. Työkalun avulla kootaan pisteistä aika-avaruuskuutiot, joka on kuvattu kuvassa oikealla kolmiulotteisena ja kaksiulotteisena. (Esri Inc, 2019.)*

Aika-avaruuskuutioista voidaan tehdä trendianalyysijä, missä seurataan kasvaako vai laskeeko pisteiden määrä eri sijainneissa tarkasteltavan ajanjakson sisällä. Tässä analyysissä käytetään Mann-Kendall trenditestä. Mann-Kendall trenditesti suoritetaan sijainnille, missä on tarkasteltavia pisteitä, ja se suoritetaan itsenäisenä yhden sijainnin aikasarjana. Kuvassa (Kuva 8) näkyy yhden sijainnin aikasarja. Mann-Kendall statistiikka on korrelaatioasioitusanalyysi kuutioiden pistemäärälle tai arvolle ja aikasarjalle. Aikasarjan ensimmäisen kuution arvoa verrataan toisen kuution arvoon. Jos ensimmäisen arvo on pienempi kuin toisen, tulos on +1. Jos ensimmäisen arvo on suurempi kuin toisen, tulos on -1. Jos arvot ovat samat, tulos on nolla. Kaikki yhden sijainnin aikasarjan vertailujen tulokset summataan yhteen. Summan odotusarvo on nolla, mikä indikoi, ettei aikasarjalla ole muutostrendiä. Kuutioiden arvojen varianssia, tasatuloksien määrää ja ajanjaksojen määrää verrataan odotusarvoon, mistä määritellään, onko muutos tilastollisesti merkittävä vai ei. Jokaisen sijainnin aikasarjan muutostrendi tallennetaan z-arvona ja p-arvona. Pieni p-arvo viittaa siihen, että trendi on tilastollisesti merkittävä. Positiivinen z-arvo viittaa nousevaan trendiin ja negatiivinen z-arvo viittaa laskevaan trendiin. Yksi visualisointimenetelmä on visualisoida laskevat ja nousevat aikasarjasijainnit eri väreillä ja värin intensiivisyys voi kertoa muutostrendin voimakkuuden. (Hamed, 2009.)



*Kuva 8: Kuvaus Mann-Kendal trenditestistä. Bin time series suomennettuna: Yhden sijainnin aikasarja. (Esri Inc. 2019)*

## 3 Tutkimusaineisto ja tutkimusmenetelmät

### 3.1 Tutkimusaineisto

Porin kaupunki käyttää WorkForce for ArcGIS -sovellusta työtehtävien jakamiseen ja seuraamiseen. Sovelluksen avulla myös historiatieto on tallessa ja tietoja käytetään tässä työssä. Tietoa on kerätty 10.5.2017 kello 09:32:06 alkaen. Tutkimuksessa käytettävällä historiatiedolla voidaan simuloida reaaliaikaista tiedonkeruuta ohjelmiston avulla. Analyysyjä varten on otettu aineistokopio, missä aineiston viimeinen päiväys on 15.12.2018 23:59:59. Taulukossa (Taulukko 2) on jaoteltu aineisto ja esitetty niiden muodot, määrät ja muut lisätietoja.

Nykyiset huoltotyön seurantatiedot:

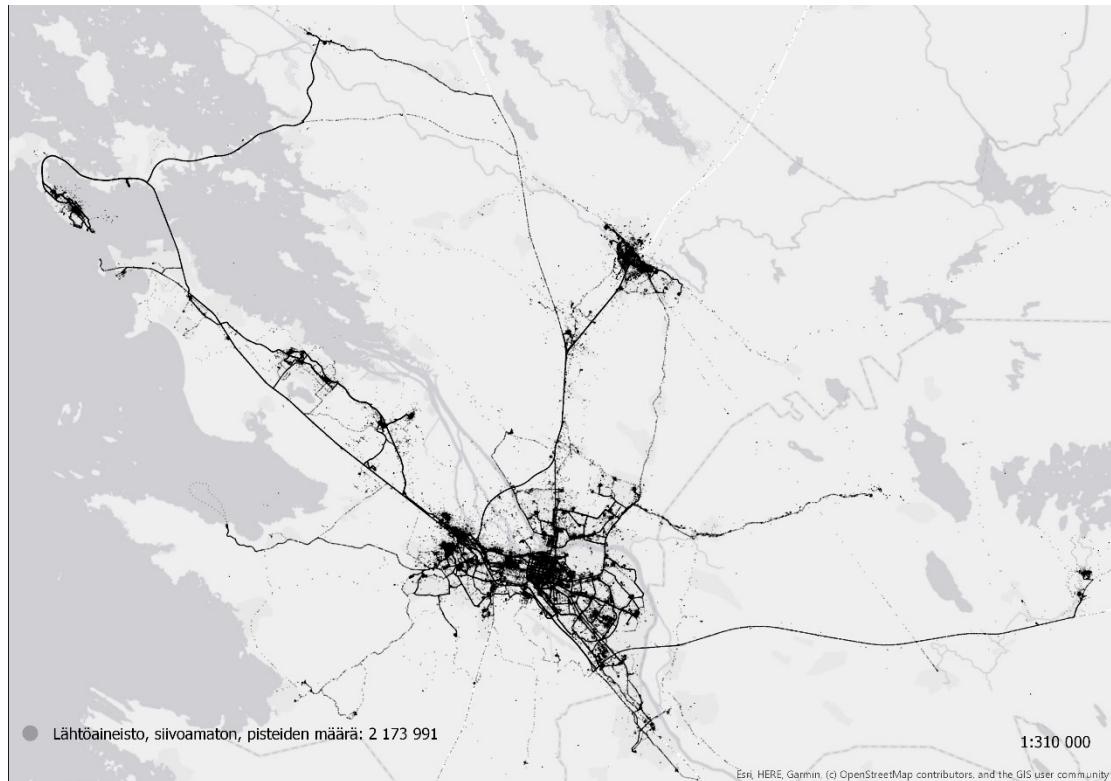
1. Työntekijät: missä ja milloin, reaaliaikainen seuranta ja keskimäärin minuutin välein jää tietokantaan merkintä sijainnista, jota voi käyttää historiatietona
2. Tehtävänannot: sijainti, tyyppi, tekijä, milloin aloitettu, milloin suoritettu
3. Kiinteistöt: kaikki Porin kaupungin huollossa olevat kiinteistöt

*Taulukko 2: Käytössä olevat aineistot*

Aineisto	Aineiston muoto	Määrä, ajankohta	Lisätietoja
Työntekijän tiedot ja reaaliaikainen sijainti (Workers)	Hosted Feature Layer, pisteaineisto, ArcGIS Online	10.5.2017 09:32:06 - tähän päivään asti. Pisteitä yhtä paljon kuin työntekijöitä (32kpl)	Käytetään vain GeoEvent Serverillä
Työntekijöiden sijainnin arkistointi noin minuutin välein (Location Tracking)	Hosted Feature Layer, pisteaineisto, ArcGIS Online	10.5.2017 09:32:06 - tähän päivään asti. 15.12.2018 mennessä pisteitä on kertynyt yli 2 miljoonaa. Kuva 9.	Aineistosta on suodatettu pois yli 100 m tarkkuudella olevat pisteet. Pisteillä ei ole vakiokeräysaikaläilyä.
Työtehtävät (Assignments)	Hosted Feature Layer, pisteaineisto, ArcGIS Online	10.5.2017 09:32:06 - tähän päivään asti.	Työtehtävä tietää milloin ja kuka suorittaa tehtävän
Kaupungin kiinteistöt	Hosted Feature Layer, alueaineisto, ArcGIS Online	411 kiinteistöä, 15.12.2018 otettu ajantasainen kopio.	Ainoastaan kiinteistöjä, jotka ovat kaupungin omistuksessa.

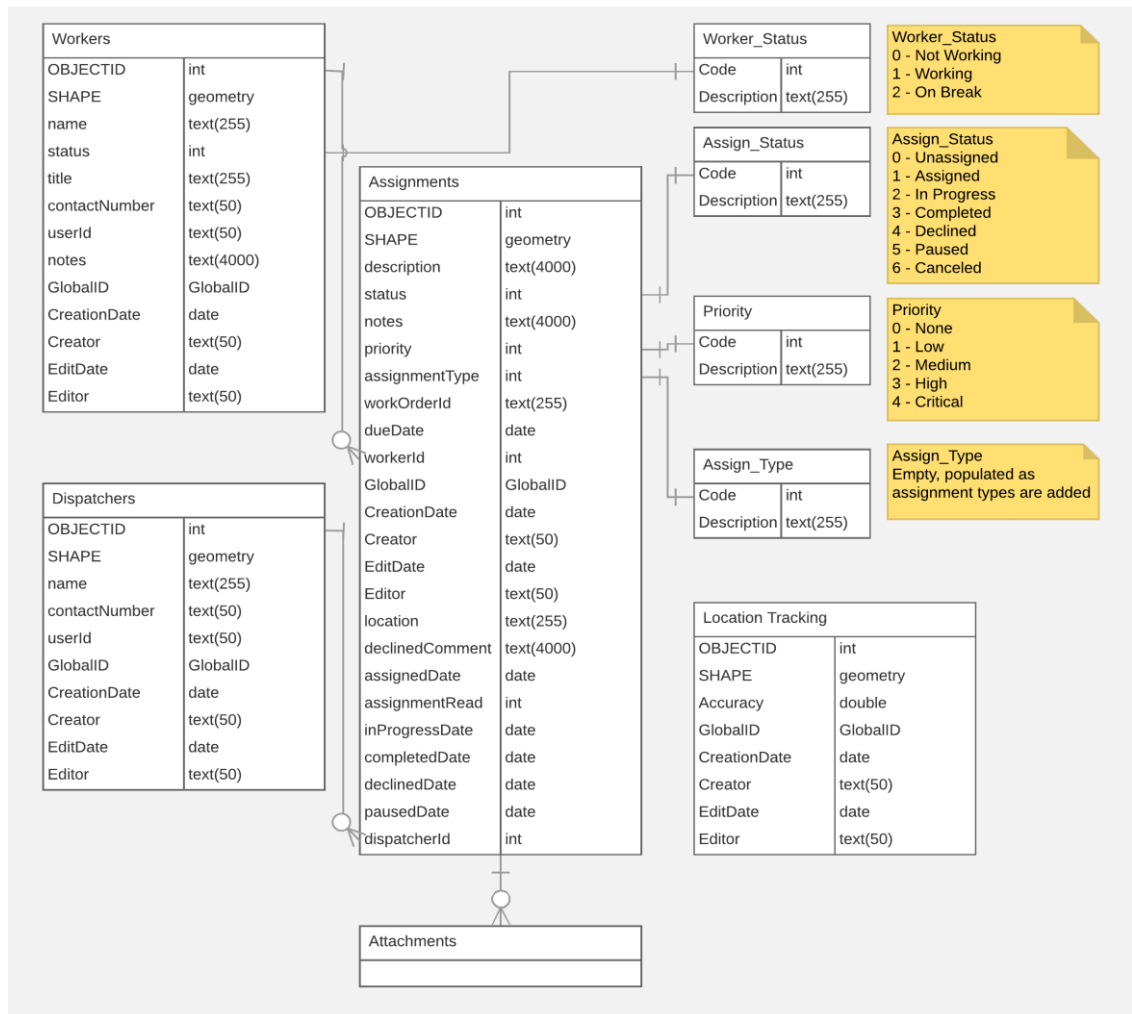
Kartta (Kuva 9) esittää lähtöaineistoa, joka on työntekijöiden sijainnin arkistointiaineisto, eli Location Tracking -pisteaineisto ennen aineiston esikäsittelyä ja pisteitä on yli kaksi miljoonaa.





*Kuva 9: Lähtöaineisto, Location Tracking -niminen pisteaineisto.*

Kaaviossa (Kuva 10) on kuvaus lähtöaineiston tietomallista ja ominaisuustietorakenteesta. Eniten käytetty aineisto on Location Tracking, joka on historiatietoa työntekijöiden sijainnista, joka on esitetty kartalla (Kuva 9). Lisäksi on hyödynnetty tehtävänantoja, nimeltään Assignments. Tietomalli on Esri Inc:n rakentama ja tarjoama Workforce for ArcGIS -sovelluksen yhteydessä. Tietomallikuvauksessa on kaikki ominaisuustiedot, jotka ovat saatavilla tähän työhön. Näiden tietojen pohjalta voi täydentää muita tietoja.



Kuva 10: Workforce for ArcGIS käytettävien aineistojen tietokantarakenne ja ominaisuustiedot. (Esri, 2018)

Aineistojen ominaisuustiedot ovat erittäin kriittisiä aineiston analysointia ja hyödyntämistä varten. Tärkeimmät ominaisuustiedot ovat:

- Työntekijän tunnusnimi (Location Tracking: Creator, Workers: name)
- Työntekijän sijainnin sijaintitarkkuus (Location Tracking: Accuracy)
- Työntekijän sijaintihistorian pisteen keräysajankohta (Location Tracking: CreationDate)
- Työtehtävän tyyppi (Assignments: assignmentType)
- Työtehtävän tekijä (Assignments: workerid)
- Työtehtävän suorittamisen ajanhetket (Assignments: inProgressDate)
- Kiinteistön kiinteistötunnus (erillinen aineisto, Kiinteistöt: Kiinteistötunnus)

### 3.2 Tutkimusmenetelmät

Työssä sovelletaan konstruktivistista tutkimusmenetelmää, jossa pyritään ratkaisemaan kiinteistöhuollon automaattinen viestintä sekä työnseuranta ajan, paikan ja työntekijän mukaan ja aineiston louhinta mahdollisimman laadukkaasti. Tutkimusmetodina on kirjallisuuskatsaus, jonka pohjalta pyritään tekemään perustellut valinnat käytettävien metodien osalta. Empiirisessä tutkimuksessa tutkimusstrategia on tapaustutkimus, joka sisältää projektityön aineiston hyödyntämisen Porin kaupungin tavoitteiden mukaisesti,

aineiston louhinnan kirjallisuuskatsauksessa löytyneillä algoritmeilla ja tutkimuksen niiden hyödyntämisestä esittelemällä ne aineiston omistajalle, Porin kaupungille.

### 3.2.1 Tutkimuksessa käytettävät ohjelmistot

Porin kaupunki käyttää paikkatiedon ohjelmistona ArcGIS-paikkatietoalustaa ja tässä tutkimuksessa käytetään tähän tuotejoukkoon kuuluvia ohjelmistoja. Tuoteisiin kuuluu valtava määrä tuotteita erilaisiin tarkoituksiin. ArcGIS-paikkatietoalusta on Esri Inc:n kehittämä ja omistama kaupallinen ohjelmistokokonaisuus. Tässä tutkimuksessa käytetään seuraavia tuotteita:

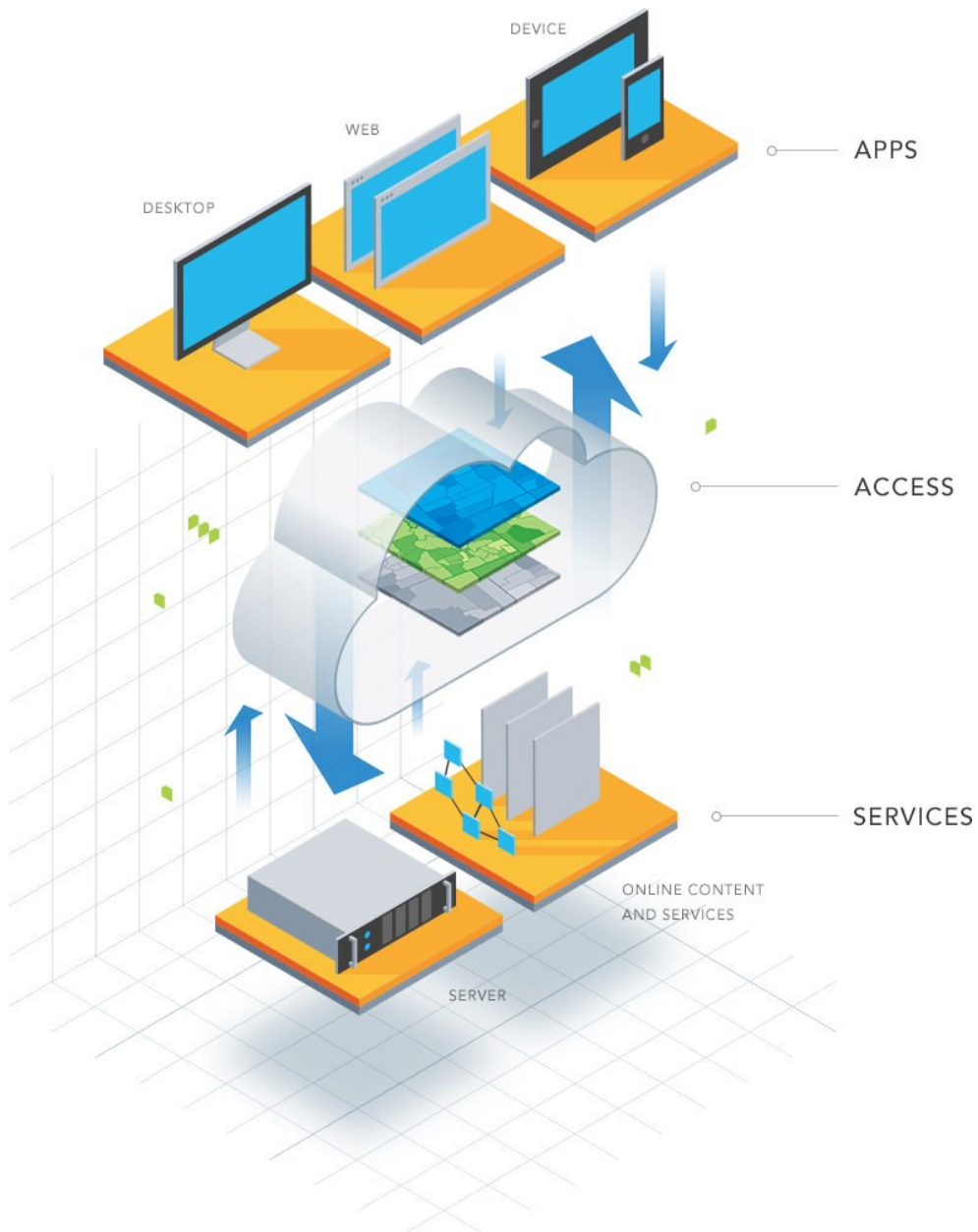
- ArcGIS Pro 2.2. (asennettava tietokoneohjelmisto) (Esri, ArcGIS Pro. 2019)
- GeoEvent Server for ArcGIS (reaaliaikaisen datan prosessointipalvelin) (Esri, ArcGIS Server. 2019)
- ArcGIS Online ja sen mahdollistamat lisäsovellukset (SaaS-palvelu, web-karttapalvelualusta) (Esri, ArcGIS Online. 2019):
  - Operation Dashboard for ArcGIS (selainpohjainen sovelluspohja) (Esri, Operation Dashboard for ArcGIS. 2019)
  - Workforce for ArcGIS (selainpohjainen sovellus ja mobiilisovellus) (Esri, Workforce for ArcGIS. 2019)

Tässä tapaustutkimuksessa käytetään ArcGIS Pro:ta aineiston esikäsittelyyn, analysoimiseen sekä visualisoimiseen kaksi- ja kolmiulotteisesti. ArcGIS Pro:sta on käytössä versio 2.2. ArcGIS Pro -ohjelmasta löytyy suurin osa työkaluista ja algoritmeista valmistyökaluina. Suurimmat rajoitteet ovat algoritmien puolella, missä ei voi käyttää klusterointiin tarkoitettuja algoritmeja viivakohteille. Tämän takia, tämän tutkimuksen analyysit rajoittuvat osin pistemäisen aineiston käsittelyyn, sillä Porin kaupungilla ei ole käytössä muita työkaluja tähän tapaustutkimukseen.

Lisäksi tässä tapaustutkimuksessa hyödynnetään reaaliaikaista aineistoa GeoEvent Server for ArcGIS -palvelimen avulla. Tällä palvelimella hyödynnetään kiinteistöhuollon aineistoa ja automatisoidaan tiedonkulku asianomaisille.

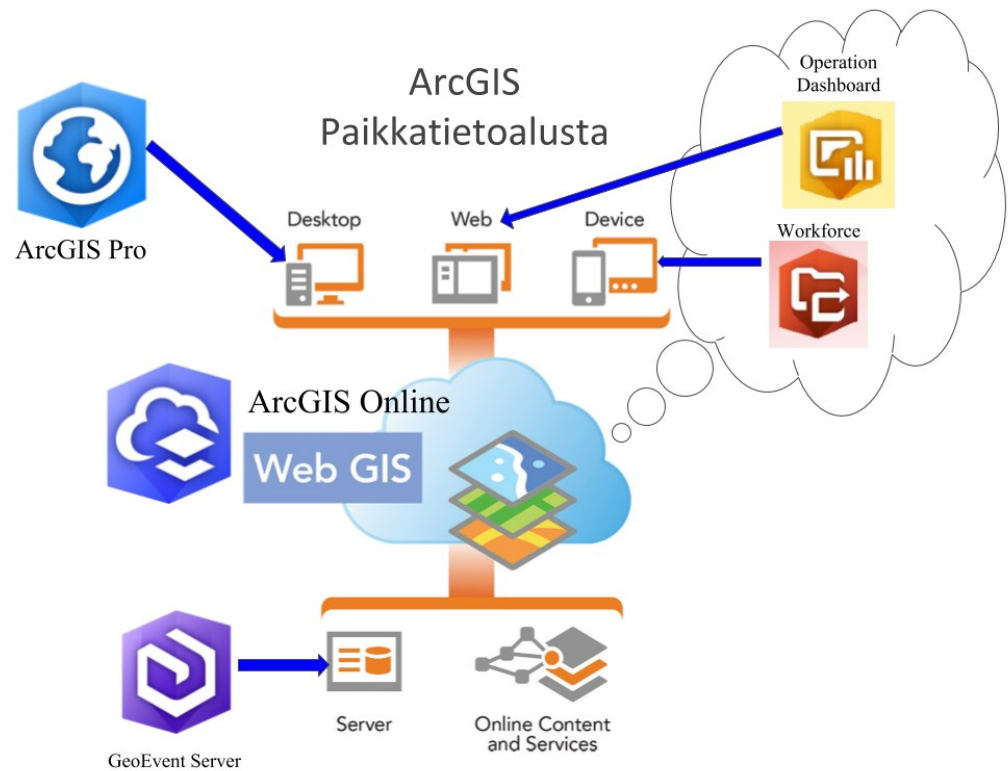
Aineisto tähän tutkimukseen kerätään mobiililaitteilla, jonka mahdollistaa älypuhelimeen asennettava Workforce for ArcGIS -sovellus. Tässä sovelluksessa on määriteltty työntekijät ja työtehtävät kartalle, joiden valmistumista ja sijaintia voidaan seurata lähes reaaliaikaisesti. Aineiston paikannus on riippuvainen työntekijän mobiililaitteesta, joka on älypuhelin ilman erillistä GNSS-paikannuslaitetta.

Jotta aineisto olisi käytettävissä myös muissa tuotteissa ja ihmisillä, niin se on tallennettu ArcGIS Online -pilvipalveluun, joka on SaaS-ohjelmisto, jota käytetään selaimen välityksellä. ArcGIS Online mahdollistaa aineiston säilytyksen, mutta myös analysoimisen ja visualisoimisen, niin kartografisilla menetelmillä kuin informaatiotuotteiden avulla, sovellukseksi asti. Tässä tutkimuksessa käytetään yhtä sovelluspohjaa: Operation Dashboard for ArcGIS.



*Kuva 11: ArcGIS Paikkatietoalustan kuvaus (Esri Deutschland, 2019.)*

Kuvassa (Kuva 11) näkyy Esrin (Esri Deutschland, 2019) kuvaus heidän paikkatietoalustastansa. Tässä tapaustutkimuksessa käytetään laajasti paikkatietoalustaa hyödyksi. Kaikki aineisto on tallennettu ArcGIS Onlineen (Access), joka mahdollistaa aineistoon pääsyn ja käytön kaikilla laitteilla (Devices). Lisäksi alimpana on GeoEvent Server (Services), joka käsittelee aineistoa.



*Kuva 12: Muokattu ArcGIS-paikkatietoalustan kuvaus (Esri, 2019)*

Käytettävät ohjelmistot on lisätty paikkatietoalustan kuvaukseen kuvassa (Kuva 12). Kuvassa on esitetty Operation Dashboard ja Workforce samassa ryhmässä, sillä ne ovat ArcGIS Onlinen (Web GIS) mahdollistamia ja ylläpidettäviä sovelluksia. Niiden käyttö ja konfigurointi tapahtuu ArcGIS Online -ohjelman kautta. ArcGIS Online on kaiken keskellä, sillä se hallitsee kaiken aineiston jaon niin maastosta huoltotyöntekijältä GeoEvent Server:lle automatisoidulle prosessille kuin ArcGIS Pro -analyysille ja informaatiotuotteille, jotka ovat tehty Operation Dashboard -sovelluspohjalla.

## 4 Tutkimustulokset

Tässä luvussa esitellään tapaustutkimuksen tutkimustulokset.

Tapaustutkimus tehtiin Porin kaupungille ja sen tavoite oli automatisoida tekstiviestilmoitukset huoltotoimenpiteen ilmoittajalle sekä löytää miten louhitaan spatio-temporaalista aineistoa erilaisilla menetelmillä niin, että aineiston hyödyllisyys on mahdollisimman suuri.

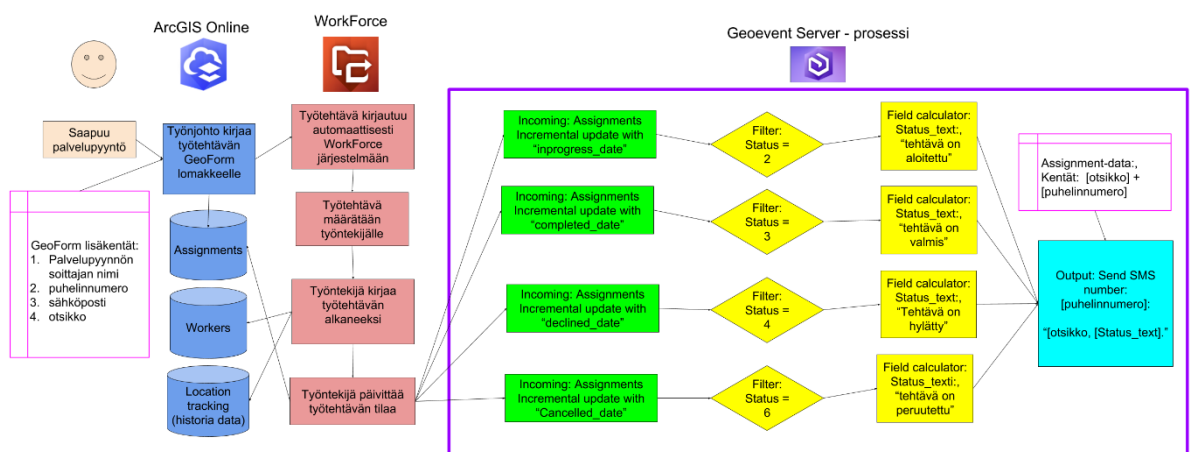
Porin kaupunki käyttää ArcGIS-ohjelmistoja aineiston keruuseen, analysoimiseen ja tallentamiseen. Aineistoa on kerätty työntekijöiltä heidän antamalla suostumuksella 10.5.2017 alkaen. Tapaustutkimuksessa käytettiin Workforce-sovelluksessa kertynyttä aineistoa hyödyksi mahdollisimman paljon, lain sallimissa rajoissa.

Kiinteistönhuollossa tehtiin töiden organisoinnin muutos loppuvuodesta 2017. Tästä johtuvat muutokset työnteossa haluttiin mitata ja se toteutettiin vertailemalla marraskuun 2017 ja marraskuu 2018 tuloksia.

### 4.1 Tapaustutkimuksen ensimmäinen tutkimuskysymys

Porin kaupungin toiveet aineiston ensisijaisesta hyödyntämisestä:  
Saada automaattiset tekstiviestit palvelupyynnön lähettäjälle.

Pyydetty käytettävä järjestelmä on GeoEvent Server for ArcGIS. Alla olevassa kaaviossa (Kuva 13 ja Liite 1), näkyy prosessikaavio tiedon keruusta tekstiviestin lähettämiseen. Prosessi perustuu tehtävänannon tilan päivittymiseen. Joka kerta kun työntekijä päivittää työtehtävän tilaa, järjestelmä kirjaa aikaleiman kyseisen tilan aikaleimakenttään. Kun aikaleimakenttä päivittyy, prosessi saa käskyn päivittää tilatekstikentän kertomaan työtehtävän tilan, ja lähettää tekstin palvelupyynnön pyytäjälle.



Kuva 13: Prosessikaavio automaattisen tekstiviestin lähettämiseen. Isompi kuva löytyy liitteestä 1.

Jotta prosessimalli toimii, Porin kaupunki on muutettava nykyisessä työnkulussa seuraavat asiat:

- Työtehtävä-aineistoon lisätään 3-5 uutta kenttää:
  - Palvelupyynnön lähettäjän nimi
  - Palvelupyynnön lähettäjän puhelinnumero
  - Tilatekstikentän, mihin annetaan tilan numeroa vastaava teksti
  - Mahdollisesti palvelupyynnön lähettäjän sähköposti
  - Mahdollisesti palvelupyynnön otsikko
- Palvelupyynnön kirjaaja ottaa käyttöön GeoForm-sovelluksen (ArcGIS Online -sovelluspohja), missä nämä uudet kenttätiedot voidaan syöttää samalla kun kirjataan työtehtävä.
  - Palvelupyynnön kirjaaja voi käyttää samanaikaisesti myös Workforce-sovellusta esimerkiksi työtehtävien jakamiseen työntekijöille.
- työntekijöitä tulee ohjeistaa tulevasta muutoksesta
  - Heidän kirjauksensa voidaan sisällyttää tekstiviestiin, joka lähtee palvelupyynnön pyytäjälle

## **4.2 Tiedon löytämisen prosessi liikeratojen hyödyntämisessä**

Työn toinen tutkimuskysymys on löytää uusia tapoja hyödyntää jo kerättyä aineistoa. Tähän tavoitteeseen pääsemiseen käytetään tiedon löytämisen prosessia. Tiedon löytämisen prosessissa (KDD) on yhdeksän vaihetta, jotka esiteltiin luvussa 2.1. Tässä luvussa tutkitaan KDD vaiheiden avulla olemassa olevaa aineistoa ja etsitään sopivia menetelmiä, joita voidaan hyödyntää kaupungin kiinteistöhuollon toiminnassa.

Ensimmäisenä määritellään loppukäyttäjän tavoitteet ja ympäristö, missä tiedon löytäminen tapahtuu. Loppukäyttäjänä toimii Porin kaupungin kiinteistöhuollon yksikkö. Kiinteistöhuollossa seurataan työntekijöiden sijaintia ja hyödynnetään sitä työtehtävien jakamisessa ja edistysen seuraamisessa. Aineistot ja tiedot ovat ArcGIS-paikkatietoalustassa reaaliaikaisessa päivityksessä ja jokapäiväisessä käytössä. Tavoitteena on automatisoida tiedonkulku työtehtävien raportoinnissa tekstiviestien avulla. Lisäksi tavoitteena on seurata ja tehostaa työntekijöiden työajan käyttöä sekä seurata Porin kaupungin kiinteistöjen huoltotoimenpiteisiin kuluvia työtunteja. Tietoja tulisi pystyä seuraamaan dynaamisella kartalla hyödyntämällä ArcGIS-paikkatietoalustaa ja mahdollisuuksien mukaan louhia tietoa reaaliaikaisesti.

Kun tavoitteet ovat selvillä, seuraavana vaiheena on kerätä ja valita aineistoja, jotka tukevat tavoitteita. Aineistot löytyvät jo valmiina ja sisältää työntekijöiden sijainnit noin minuutin tarkkuudella, huollon työtehtävät sekä huollettavat kiinteistöt. Tietoa aineistosta tarkemmin löytyy luvussa 3.1.

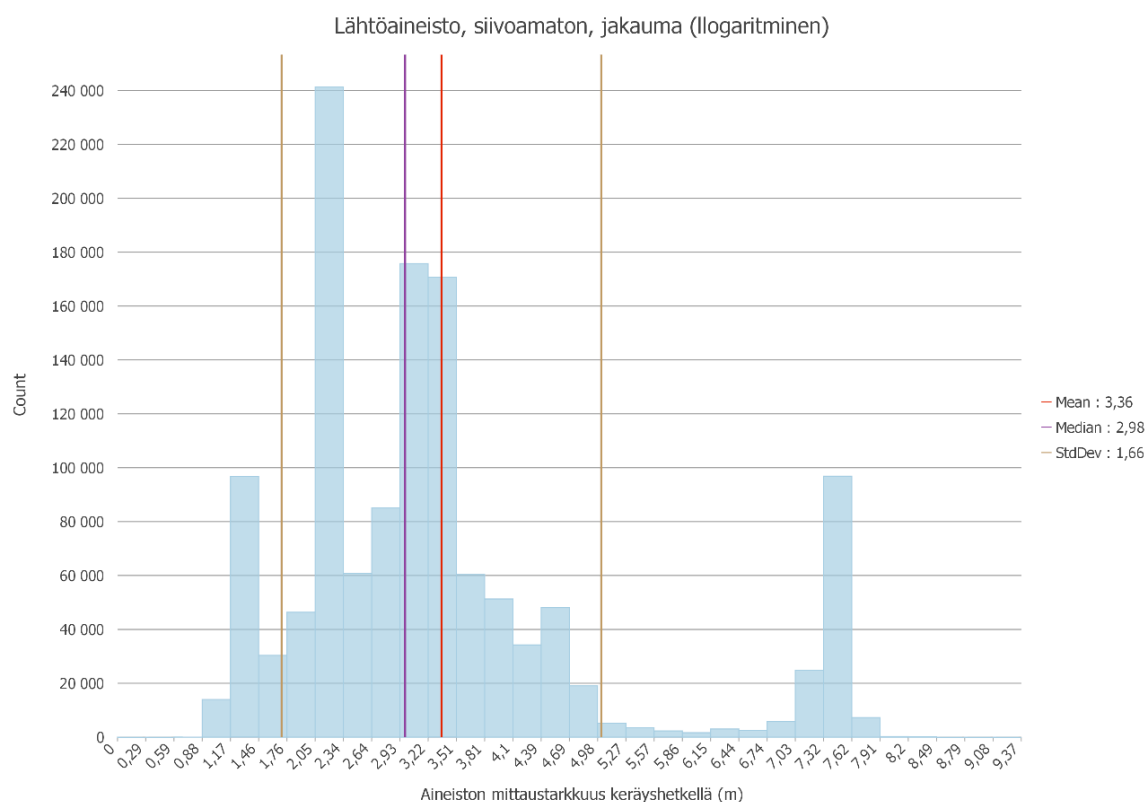
Kun aineisto on hankittu, seuraava vaihe on aineiston käsittely, jotta tutkittavan aineiston laatu on soveltuvaa käyttötarkoitukseensa. Aineiston ominaisuustietoja muokattiin sopimaan käyttötarkoituksia. Myös neljäs vaihe, aineiston muuntaminen, toteutui samaan aikaan, sillä aineistosta otettiin paikallinen kopio, johon halutut muutokset tehtiin. Paikallisessa kopiassa on alusta 15.12.2018 asti kerätyt tiedot. Aineiston käsittely ja muuntaminen tapahtui iteroimalla. Ensin tarkasteltiin lähtöaineistoja, sitten huomattiin muutostarpeet ja muunnettiin aineisto sopivaan muotoon ja säilytyspaikkaan toteutusta varten, jatkettiin aineiston käsittelyä ja huomattiin lisää muutostarpeita

aineistoanalyysivaiheessa. Muutoksia ei voida tehdä suoraan alkuperäiseen aineistoon, sillä se on jo tuotannossa huoltotoimenpiteiden työnohjauksessa. Näissä muutoksissa on otettava huomioon, miten ne voidaan toteuttaa myös tuotantoversioon, ettei viimeisessä vaiheessa (9) tule peruuttamattomia ongelmia.

Lähtöaineiston tarkkuutta tutkitaan ”viiden mainekkaan” kriteerin pohjalta (Taulukko 3) (Guptill ja Morrison 1995).

*Taulukko 3: Lähtöaineiston tarkkuuden tarkastelu*

Kriteeri	Lähtöaineiston tarkkuus kriteerin valossa
Sijaintitarkkuus	Location Tracking -aineistossa on ”Accuracy (m)” -kenttä, mistä saadaan sijaintitarkkuus (Kuva 10). Histogrammit lähtöaineistosta alkuperäisenä (Kuva 14) ja siivottuna (Kuva 15).
Ominaisuustietotarkkuus	Luokiteltavissa kentissä on käytetty arvoluetteloita, joten niiden tarkkuus on riittävällä tasolla (Kuva 10).
Temporaalinen tarkkuus	Aineistoa on kerätty sekunnin tarkkuudella, joka on riittävä tähän työhön.
Looginen eheys	Aineiston looginen eheys on hyvä, sillä se vastaa annettua tietomallia (Kuva 10).
Valmius ja täydellisyys	Tietomallikuvaus ja dokumentaatio ovat ajan tasalla ja vastaa täysin lähtöaineiston tietomallia (Kuva 10). (Veregin 1998)



*Kuva 14: Lähtöaineiston mittaustarkkuus*



Tietoja, joita ei ollut valmiina, jotka laskettiin aineistoon:

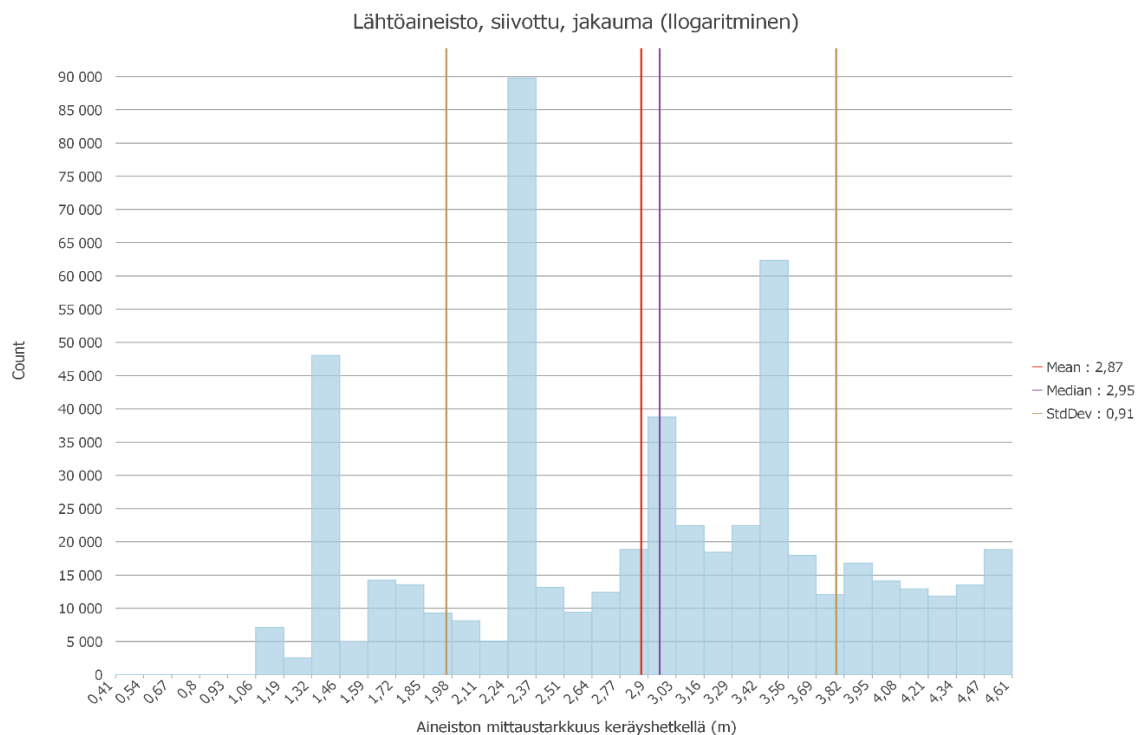
- Työntekijän tunnusnimi puuttuu tehtävänannosta, siellä on vain tunnusnumero. Yhdistävä taulukko puuttui.
  - Se luotiin hyödyntämällä työtehtävän muokkaajatietoja, millä yhdistettiin tekijä ID numeroon.
- Huoltotoimenpiteen pyytäjän yhteystiedot (puhelinnumero)
  - Luotiin uusi sarake uutta tietoa varten.
- Huoltotoimenpiteeseen liittyvä kiinteistö
  - Aineistoon lisättiin uusi sarake kiinteistötunnusta varten.
  - Vaatii GeoForm-prosessin, jos se halutaan liittää reaaliaikaisesti tehtävänantoihin.
- Työntekijän sijaintihistoriasta tieto siitä, mitkä ovat työaika
  - Lisättiin kenttä, mihin luokitellaan kellonajan mukaan työaika ja vapaa-aika.

Tunnettu puute:

- Sijaintihistoria-aineistoon ei jää merkintää työntekijän tilasta; onko hän töissä, tauolla vai vapaalla, tai mitä työtehtävää tekijä on suorittamassa.
  - Oletus on, että sovellus kerää sijaintia aktiivisesti vain ”töissä”-tilassa.

Tietoa suodatettiin seuraavalla tavalla:

- Vain alle 100 m tarkkuudella olevat sijaintihistoriatiedot hyödynnettiin analyyseissä.
- Vain arkipäivinä kello 07:00-15:00 välillä kertyneet työntekijöiden sijainnit hyödynnettiin analyyseissä.
- Käytetään vain niiden työntekijöiden sijainteja, jotka ovat koko keräysajan aikana töissä, jotta aikajaksojen vertailu on vertailukelpoista muutoksen esittämiseen.
- Kaikki pisteet, jotka ovat Porin kaupungin ulkopuolella poistettiin. Muutama piste löytyi Suomen valtion rajojen ulkopuolelta.
- Kuva 15 esittää siivotun aineiston mittaustarkkuutta.

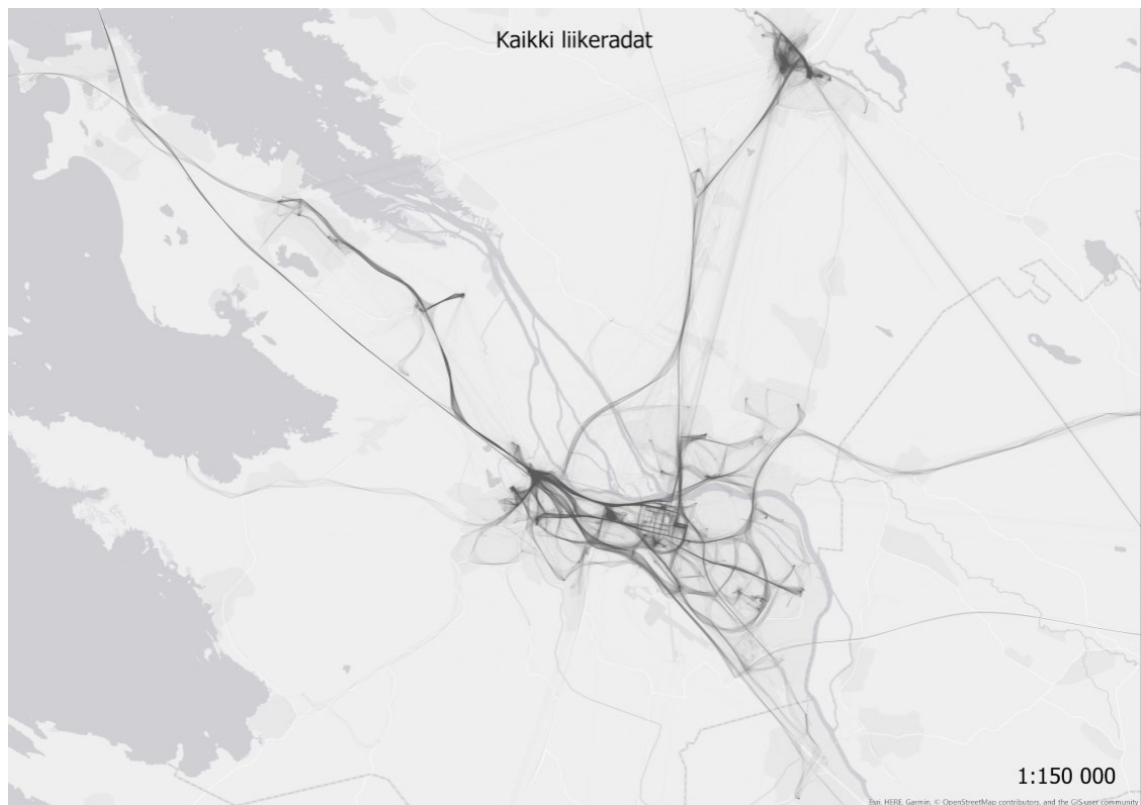


Kuva 15: Siivotun lähtöaineiston mittaustarkkuus

Kaikki liikeradat

1:150 000

© 1998, 2000 OpenStreetMap contributors, and the OpenStreetMap community



*Kuva 17: Kaikki liikeradat visualisoinnin kanssa, symboliikan referenssimittakaava on 1:1000.*



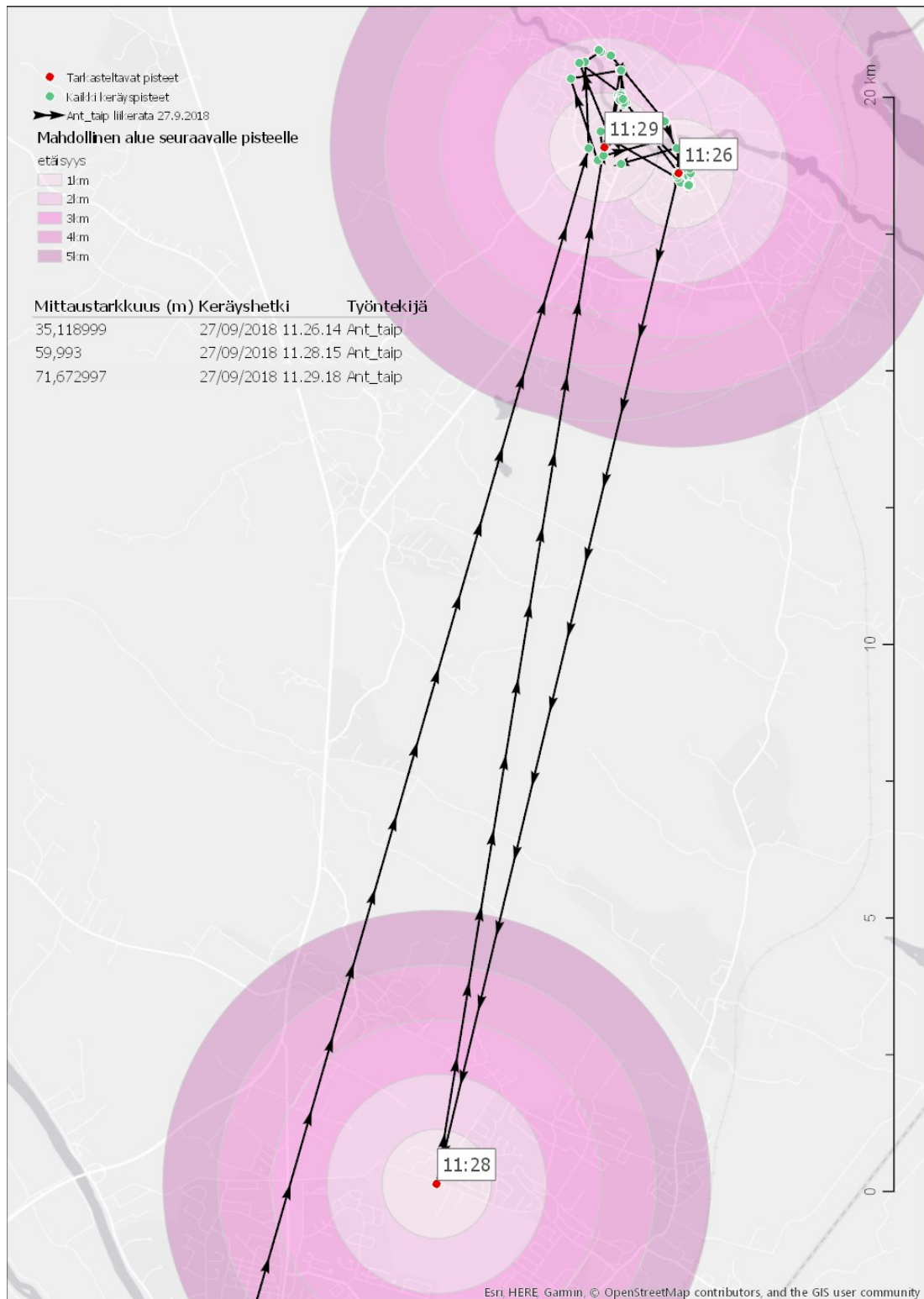
*Kuva 18: Kaikki liikeradat Porin keskustassa. Yksi viiva on yhden työntekijän yksi työpäivä. Kartassa on 1:1000 referenssimittakaava ja läpinäkyvyyttä, jotta liikeratojen suosituimmat solmukohdat ja reitit korostuvat.*



*Kuva 19: Työntekijöiden liikeradat syyskuu 2018 aikana, päiväkohtainen jako.*

Alustavan aineiston suodattamisen jälkeen aineistossa on edelleen kohinaa ja virheellistä aineistoa. Aineiston laatu on verrattain heikko, sillä aineistoa kerättiin älypuhelimilla, joiden päätehtävä ei ole olla sijaintitiedon keräyslaite ja keräysväli on parhaimmillaan melko pitkä, noin minuutin välein ja huonoimmillaan jopa 40 minuuttia. Pelekisin ja Theodoridiksen (2014) mukaan aineiston laatuun vaikuttaa suuresti sekä sen keräysmenetelmä että keräysväli. Näitä tietoja voidaan myös hyödyntää aineiston siivoamiseen, sillä niiden avulla voidaan tehdä oletuksia kyseistä aineistosta. Niiden avulla pystytään arvioimaan maksimialue, joka kahden peräkkäisen keräyspisteen väliin mahtuu. Työntekijät liikkuvat sekä kävellen että autolla, joten vaihteluväli on melko suuri. Taajama-alueella pystytään arvioimaan, että työntekijä ei aja kovempaa vauhtia kuin 60 km/h, eli kulkevat minuutin aikana maksimissaan yhden kilometrin. Taajaman ulkopuolella nopeus voi nousta jopa 100 km/h, mikä tarkoittaa keskimäärin 1,667 km kulkemista minuutissa. Pisteiden keräysväli vaihtelee aktiivisella keräysajalla 0-3 minuutin välillä, mikä tarkoittaisi maksimissaan 5 km kulkemista kerättyjen havaintojen välillä. Alla olevassa kartassa on esimerkki virhepisteestä, missä yksi piste poikkeaa huomattavasti (Kuva 20). Näiden virhepisteiden etsiminen on mahdollista, mutta näin suurella aineistolla, se vaatii paljon laskentatehoa ja ohjelmointia.

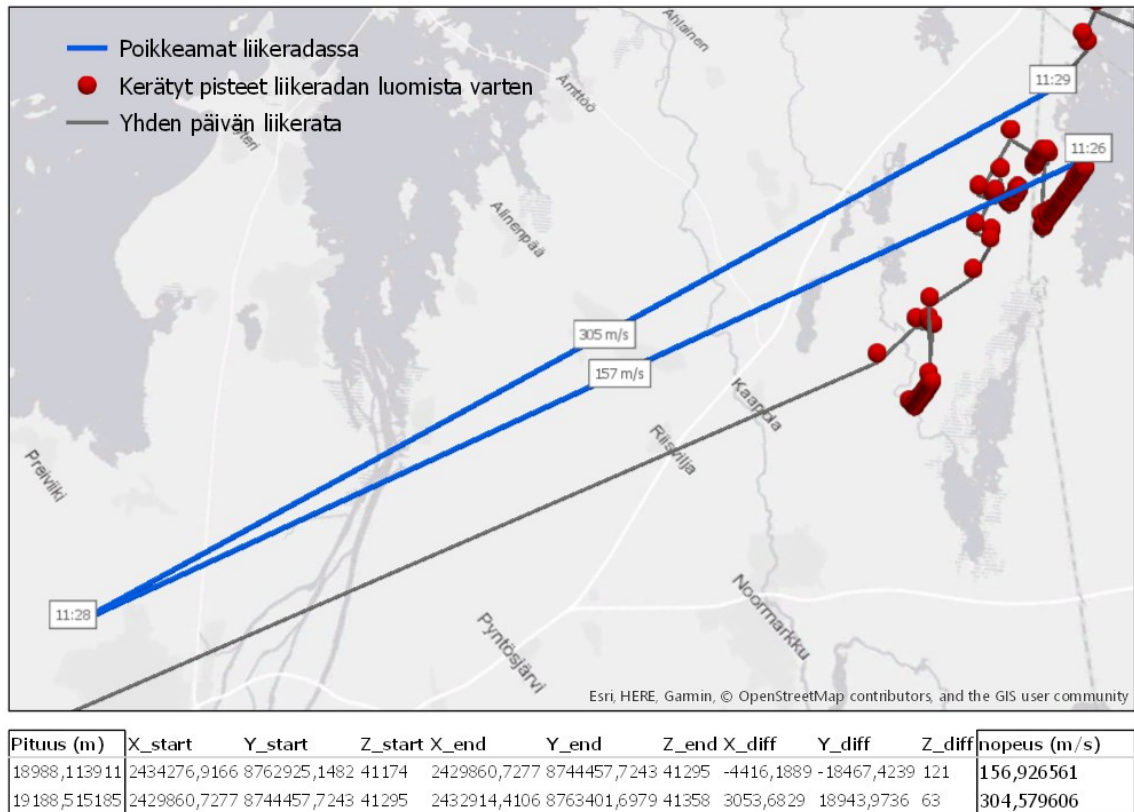




Kuva 20: Työntekijän liikerata ja maksimialue seuraavalle pisteelle

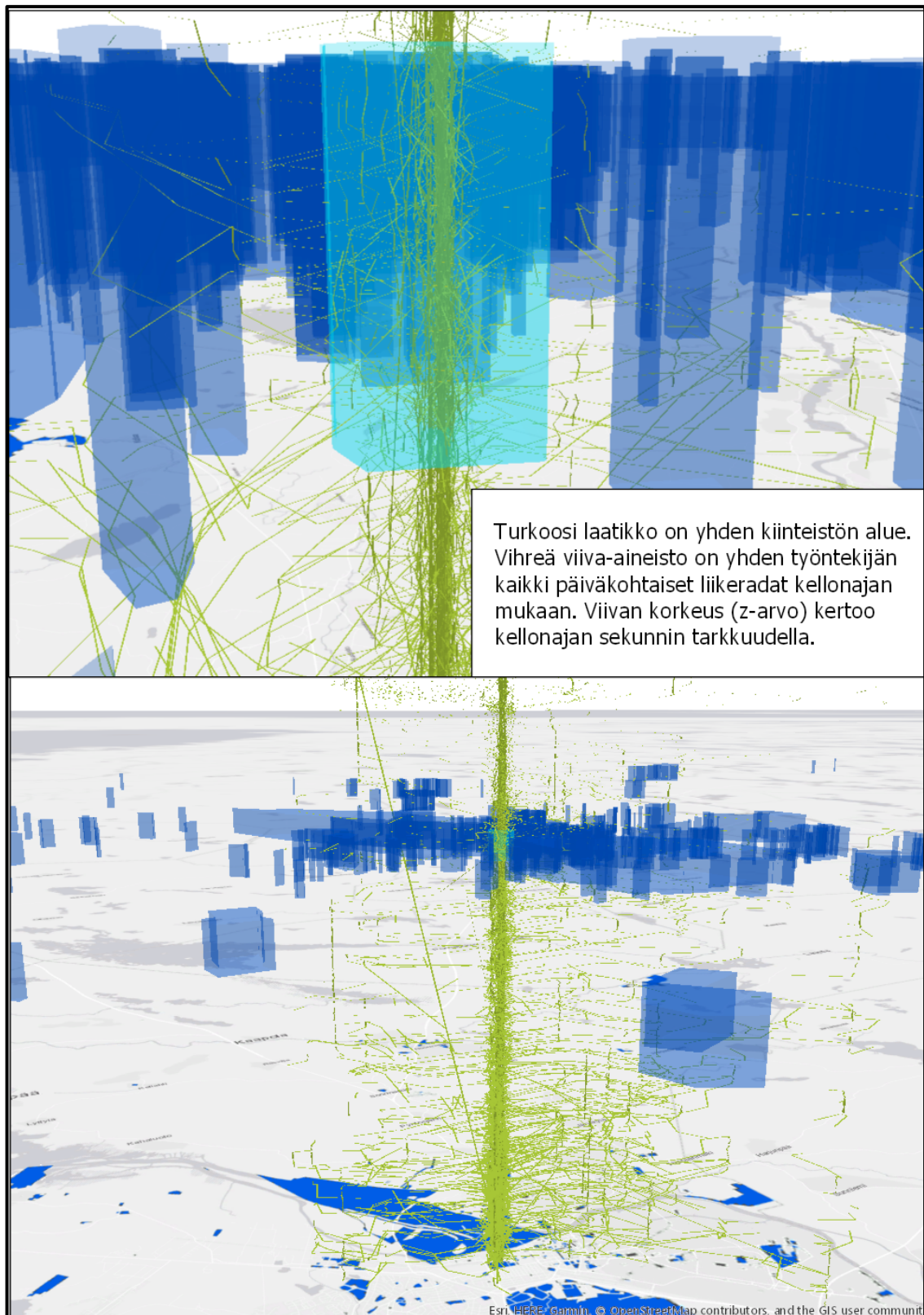
Paikkatietokanta käsittelee pisteaineiston kohteita itsenäisinä kohteina, jotka eivät riipu järjestyksestä. Tämä aiheuttaa sen, että pisteaineistolla ei pysty laskemaan puhtaasti pisteiden arvojen erotusta. Ratkaisu tähän on luoda pisteistä viiva, missä jokainen taitepiste on kerätty piste. Jotta aikaleima saadaan säilytettyä, tieto voidaan syöttää viivalle z-arvoksi. Jokainen taitepiste sisältää tiedon  $a = (x,y,z)$ , missä x ja y ovat koordinaattitiedot ja z on aika. Kun jokaisen peräkkäisen pisteen välille luodaan viiva, sen etäisyys pystytään laskemaan niin koordinaateissa kuin aikana. Matka ja aika kertovat

nopeuden, minkä perusteella pystytään laskemaan jokaisen keräysvälin matkan nopeus. Jos nopeus on liian suuri, piste on häiriöpiste. Häiriöpisteet on poistettava ja liikerata laskettava uudelleen. Kartassa (Kuva 21) näkyy selkeän häiriöpisteiden nopeusarvot, jotka ovat kulkuneuvon huomioon ottaen mahdottomia ja näin ollen häiriöpiste voidaan poistaa.



Kuva 21: Liikerata esitettynä kolmiulotteisena, missä kolmas ulottuvuus on aika. Kuvassa on korostettuna häiriöpisteet ja niiden arvot. Kuvassa näkyy myös taulukkotietona viivojen pituus, koordinaattitiedot viivojen päästä sekä niistä laskettu nopeus.

Aineistosta onnistuttiin laskemaan kaikille liikeradan välisille pisteille nopeudet ja sen perusteella suodattamaan liikeradoista häiriöpisteet pois. Tämä kuitenkin vaatii useamman iteroitokerran, sillä, kun piste poistetaan ja yhdistää jäljelle jääneet pisteet uudelleen, syntyy uusi viiva, jonka nopeutta ei tiedetä ja se tulee tarkistaa uudelleen. Aineistoa tarkastellessa huomataan, että tällaisia häiriöpisteitä on monesti 2-3 samassa ”häiriösjainnissa” ja sitten sijainti palaa takaisin alkuperäiselle sijainnilleen. Näiden häiriöiden syy ei ole tiedossa, mutta todennäköisesti johtuu älypuhelimien heikosta signaalista esimerkiksi virransäätöasetuksien takia. Kun älypuhelimien asettaa virransäätötilaan ja sammuttaa näytön, puhelin laittaa GPS-haun pois päältä hetkellisesti.



*Kuva 22: Kuvissa on sama aineisto, mutta eri katselukulmasta. Tämän kyseisen työntekijän liikeradat kohtaavat eniten samassa kiinteistössä riippumatta kellonajasta, sillä muodostuva pylväs on selvästi erottuva maasta ylös asti.*

Kartassa (Kuva 22) näkyy selkeä kuvio aineistosta, missä korkeusarvo on jokaisen työpäivän kellonaika klo 07:00 ja 15:00 välillä sekunnin tarkkuudella. Vahva pylväsmäinen kuvio viittaa siihen, että kohteessa on vierailtu usein ja mihin tahansa



kellonaikaan työpäivän aikana. Tämä toistuu usealla työntekijällä, mikä viittaa siihen, että tietyt työntekijät ovat pitkään huolehtineet tietyistä kiinteistöistä.

Pelekis ja Theodoris (2014) ehdottivat myös toista ratkaisua, viivojen tasoittamista ja yksinkertaistamista. Tulokset erilaisista työkaluista, kuten “Simplify Line”, “Smooth Line” tai “Generalize” ei tuottanut haluttuja tuloksia. Keräyspisteet, jotka olivat selvästi virheellisiä, olivat niin kaukana toisistaan, että ne muodostivat niin pitkät viivat, että algoritmi suosi juuri niiden viivojen osia.

Seuraava vaihe aineiston esikäsittelyssä on segmentoida aineisto. Zheng et al. (2010 s.8-9) menetelmä muutospistesegmentointi osoittautui toimivaksi menetelmäksi, sillä heidän tutkimuksensa osoitti, että liikeradan nopeudesta pystyy päättelemään seurattavan kohteen kulkumuodon. Tässä tapaustutkimuksessa oletetaan, että työntekijät liikkuvat autolla tai kävellen kiinteistöllä ja työtehtävissään, joten tarkasteltavia kulkumuotoja on kaksi, kävely ja autolla ajaminen. Zheng et al. (2010, s.8-9) tutkimuksen mukaan 95% kävelystä tapahtuu noin 3 m/s alapuolella ja autolla ajaminen 92% 5 m/s yläpuolella.

Kun aineiston segmentoi tällä menetelmällä kahteen kategoriaan (3-5 m/s jää kategorisoimatta), niin kartalla (Kuva 23) näkyy miten eri kulkumuodot asettuvat.



*Kuva 23: Kuvassa on kaikki liikeradat (suurimmat häiriöviivat poistettu), missä liikeradat on segmentoitu nopeuden mukaan kulkumuotoihin.*

Kartasta (Kuva 23) voidaan tulkita, että suurin osa kävelysegmenteistä klusteroituvat autoilusegmenttien risteyskohtiin. Nämä risteyskohdat ovat tunnettuja kiinteistöjä. Autosegmentit noudattavat tunnettua tieverkon muotoa. Tätä aineiston segmentointia voidaan hyödyntää ja jatkojalostaa aineiston luokittelussa ja klusteroinnissa.

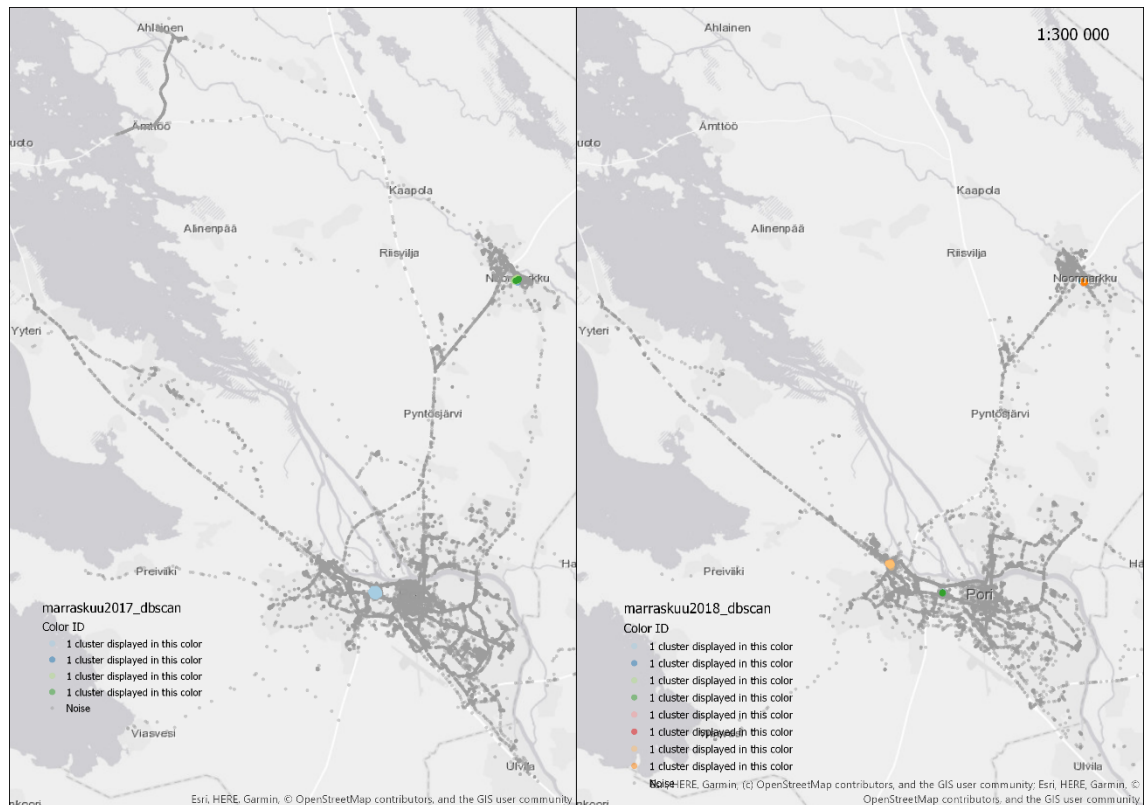


Viides vaihe on aineiston louhintamenetelmän valinta. Tämä vaihe toteutettiin tiedon löytämisen prosessin vastaisesti, sillä tavoitteena oli löytää mahdollisimman paljon erilaisia aineiston hyödyntämisen mahdollisuuksia ja vertailla tuloksien hyödyllisyyttä. Tässä vaiheessa rajataan louhintamenetelmät klusterointiin, luokitteluun ja visuaalisiin menetelmiin. Louhimismenetelmien vaiheet käsitellään seuraavissa luvuissa tarkemmin. Myös luohinnan algoritmien valinta ja ajo (vaiheet 6 ja 7) käsitellään seuraavissa luvuissa. Vaiheet 8 ja 9 käsitellään luvussa 5, missä arvioidaan tuloksia, tulkitaan ja hyödynnetään saatavaa tietoa.

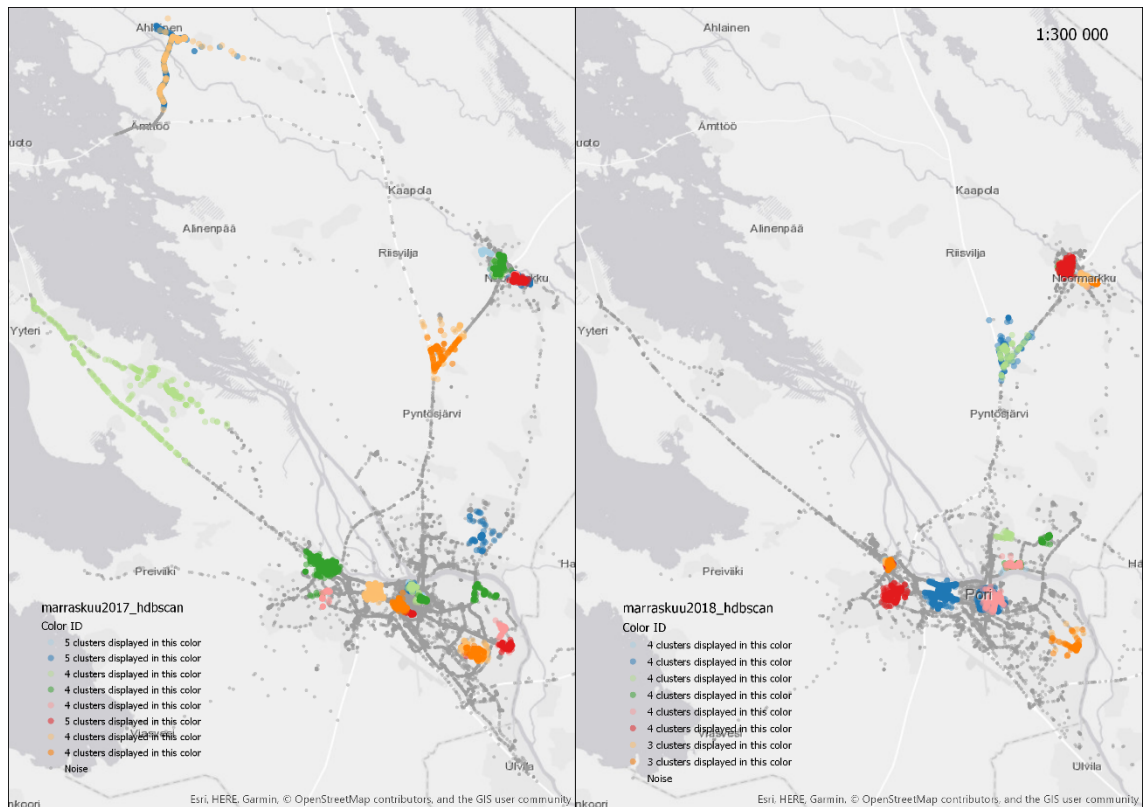
#### 4.2.1 Klusterointimenetelmien tulokset

Klusterointimenetelmäksi valittiin tiheyspohjainen klusterointialgoritmi kerätyllä pisteaineistolla. Käytettävät algoritmit ovat DBSCAN, HDBSCAN ja OPTICS. Algoritmien vertailua varten käytettiin samoja parametreja algoritmin ajamiseen, jotka eivät olleet optimaalisia kaikkiin analyysihin, mutta tuottivat tuloksia jokaisessa analyysissä. Analyysissä käytettävä pistemäärä nosti analyysin suoritusaikaa, joten aineistoa rajattiin yhden kuukauden aineistoon. Tässä menetelmässä seurataan työn organisointimuutoksia, joten algoritmit on ajettu marraskuu 2017 ja 2018 -aineistoille.

Ensimmäisenä DBSCAN-algoritmin tulokset näkyvät kuvassa (Kuva 24). Pisteiden määrä on puolittunut kuukausien välillä, mutta analyysin parametrit pidettiin samana. 2017 aineistosta löytyi huomattava määrä enemmän klustereita, mutta kuitenkin samoilta alueilta löytyi klusterit myös vuonna 2018. Klustereita löytyy myös samalta kiinteistön alueelta. Yhdeltä alueelta syntyi uusi klusteri vuoden aikana (läntisin klusteri 2018).



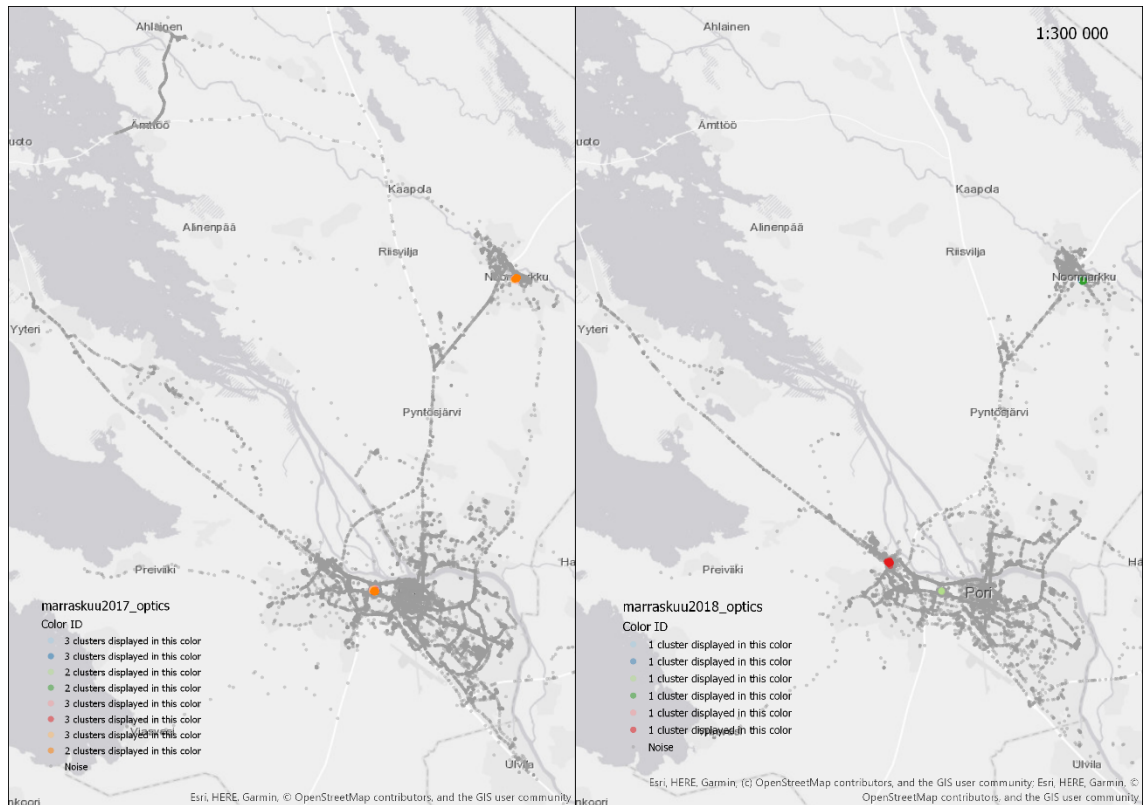
Kuva 24: DBSCAN tulokset marraskuussa 2017 ja 2018.



Kuva 25: HDBSCAN tulokset marraskuussa 2017 ja 2018.

Kuvassa (Kuva 25) näkyy HDBSCAN tulokset, jotka ovat huomattavasti erilaiset DBSCAN tuloksiin. HDBSCAN eroaa DBSCAN:sta, siten, että se optimoi itse klustereiden koon ja hakuetaisyyden aineiston perusteella ja vähemmän riippuvainen annetuista parametreissa. Klustereiden määrässä on huomattava ero, mutta myös osa klustereista on hävinnyt alueelta kokonaan. Esimerkiksi luoteessa olevassa kohteessa ei ole juurikaan pisteitä 2018. 2018 aineistossa on tehty myös laajempia klustereita, missä on vähemmän päällekkäisyyksiä.

Kuvassa (Kuva 26) on OPTICS tulokset, jotka ovat samankaltaiset DBSCAN tuloksien kanssa. Klustereiden sijainnit molemmissa analyyseissa ovat hyvin samankaltaisia, sillä OPTICS perustuu DBSCAN algoritmiin, mutta luo sen lisäksi saavutettavuustaulukon, missä hienosäädetään klustereiden sijaintia. OPTICS algoritmin ajo kesti kaikista pisimpään ja tuotti käytännössä saman tuloksen kuin DBSCAN.



Kuva 26: OPTICS tulokset marraskuu 2017 ja 2018.

Kaikissa kolmessa klusterointianalyysissä tulokset viittaavat siihen, että klustereita löytyy pääasiassa tunnettujen kiinteistöjen kohdalla. Tarkemmassa tarkastelussa esimerkiksi yksittäisten työntekijöiden keräyspisteet DBSCAN algoritmilla antaa samanlaisia tuloksia, mutta tunnetut kiinteistöt ovat hyvin erilaisia työntekijöiden välillä. Tämä viittaa siihen, että työnjakoa on tehty kiinteistöittäin, mikä vaikuttaa tuloksiin. Lisäksi tuloksissa löytyy myös muita alueita, jotka eivät ole kiinteistöjä, mutta ovat klustereita. Tulokset näistä DBSCAN analyysistä löytyy liitteessä 3.

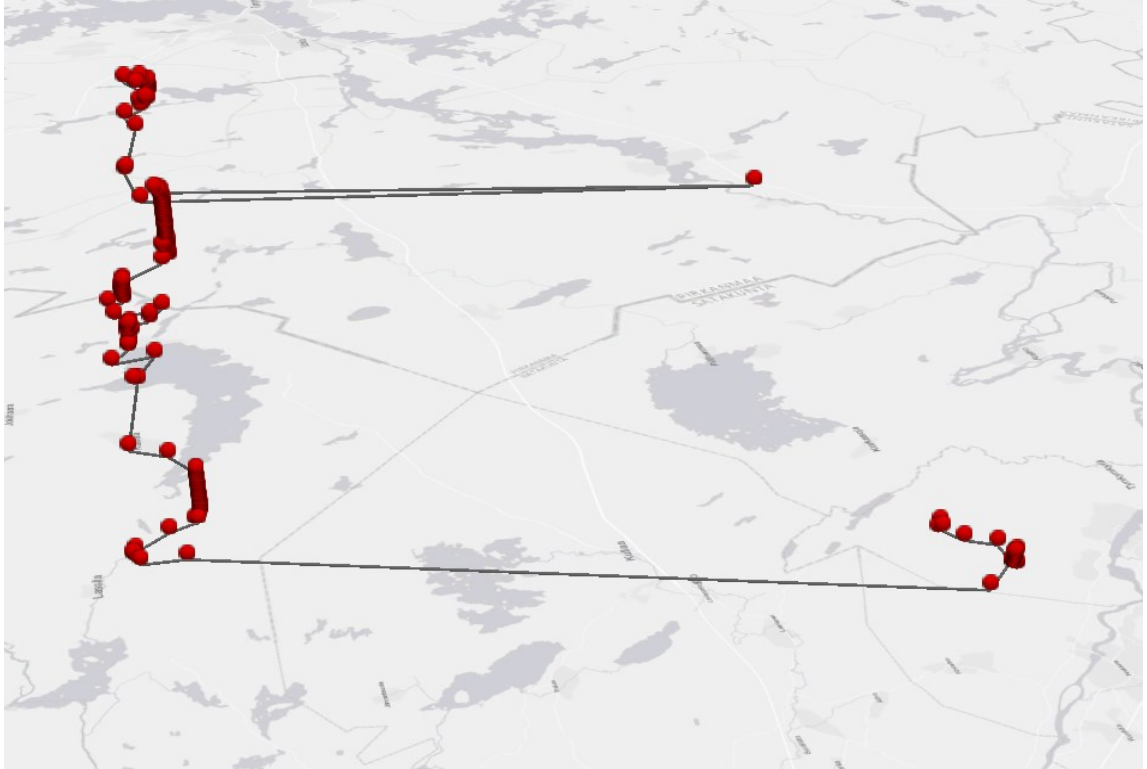
Otannan perusteella klustereita löytyy huollettavien kiinteistöjen alueilta. Kun tarkastellaan yksittäisiä työntekijöitä, niin esille tulee myös klustereita, jotka eivät ole tunnettujen kiinteistöjen alueilla (liite 3).

Vertailussa olevien ajanjaksojen, marraskuu 2017 ja marraskuu 2018, käytettävien aineistojen kokoero on huomattava: 2017 aineistossa on 68000 pistettä ja 2018 on 41000 pistettä. Tämän takia 2017 löytyi enemmän klustereita HDBSCAN ja OPTICS analyysissä, mutta DBSCAN analyysissä niitä löytyi vähemmän.

#### 4.2.2 Luokittelumenetelmien tulokset

Liikeratojen luokittelu vaatii viivojen segmentoinnin luokittelun semantiikan mukaisesti, sillä aineistoa voidaan luokitella vain viivakohtaisesti, joten viivat pitää purkaa tarkoituksen mukaisiin osioihin. Kun tarkastellaan yksittäisiä liikeratoja, huomataan käyttäytymisen eroa eri alueilla. Kiinteistöjen rajojen sisällä, liikeradat liikkuvat tiuhaan edestakaisin, mutta pysyvät rajojen sisäpuolella. Kun liikeradat ovat lähellä tieverkostoa, niin liikeradat noudattavat samoja muotoja kuin tieverkko. Kun lasketaan taitepisteiden

välisen kulkunopeuden, sitä voidaan hyödyntää viivojen segmentoinnissa. Jos nopeus on pieni, seurattava kohde on paikallaan ja kun nopeus on suuri, kohde on ajamassa autoa. Porin kaupungin kiinteistöjen huoltohenkilökunta liikkuu autolla huoltokohteesta toiseen ja siirtyy kiinteistön alueelle hoitamaan työtehtävää. Työntekijän on pysäytettävä auto, jotta hän voi hoitaa työtään, joten jos aineistosta voidaan löytää pysähtyneet kohteet, niin liikeradat pystytään luokittelemaan työntekijän tehtävän mukaisesti. Kuva 27 esittää yhtä liikerataa kolmiulotteisesti, missä punaiset pallot ovat taitepisteitä, eli lähtöaineisto. Liikerataa segmentoidaan näissä taitepisteissä, jotta sitä voidaan luokitella erikulkumuotoihin.



*Kuva 27: Kolmiulotteinen näkymä yksittäisestä liikeradasta. Punaiset pisteet ovat taitepisteitä, eli aineiston kerätyt pisteet.*

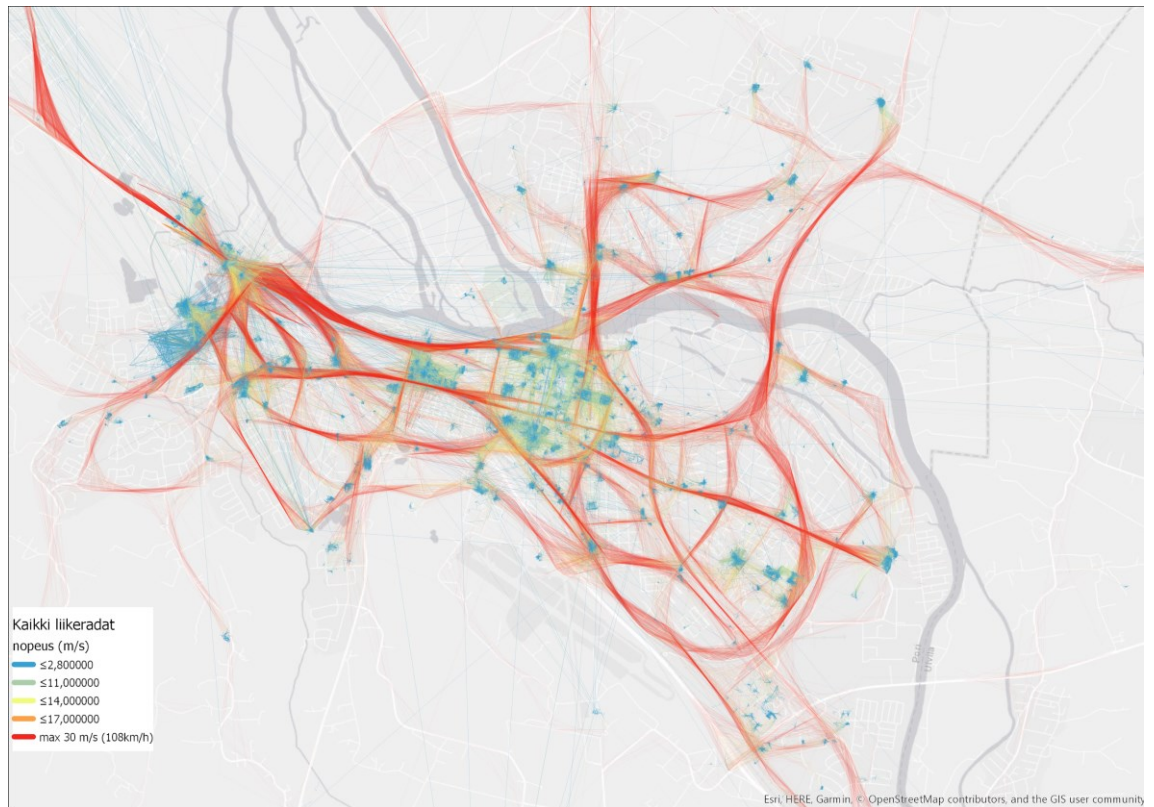
Tavoitteena on luokitella missä ja milloin työntekijä on tekemässä huoltotöitä kiinteistöllä ja milloin autossa ajamassa. Kun tunnetaan Porin kaupungin kiinteistöt ja tieverkon, liikeradat pystytään lajittelemaan näiden tunnettujen kohteiden perusteella päällekkäisyysanalyysin avulla. Toinen vaihtoehto on määritellä, mikä on maksimi nopeus kummassakin tapauksessa. Liikeradat luokiteltiin nopeuden perusteella seuraavalla tavalla:

- Kiinteistöllä: nopeus alle 2.8 m/s (noin 10 km/h)
- Kantakaupunki ajamista: nopeus alle 11 m/s (noin 40 km/h)
- Taajamassa: nopeus alle 14 m/s (noin 50 km/h)
- Taajaman läheisyydessä: nopeus alle 17 m/s (noin 60 km/h)
- Maantieajoa: nopeus maksimissaan 30 m/s (noin 108 km/h)

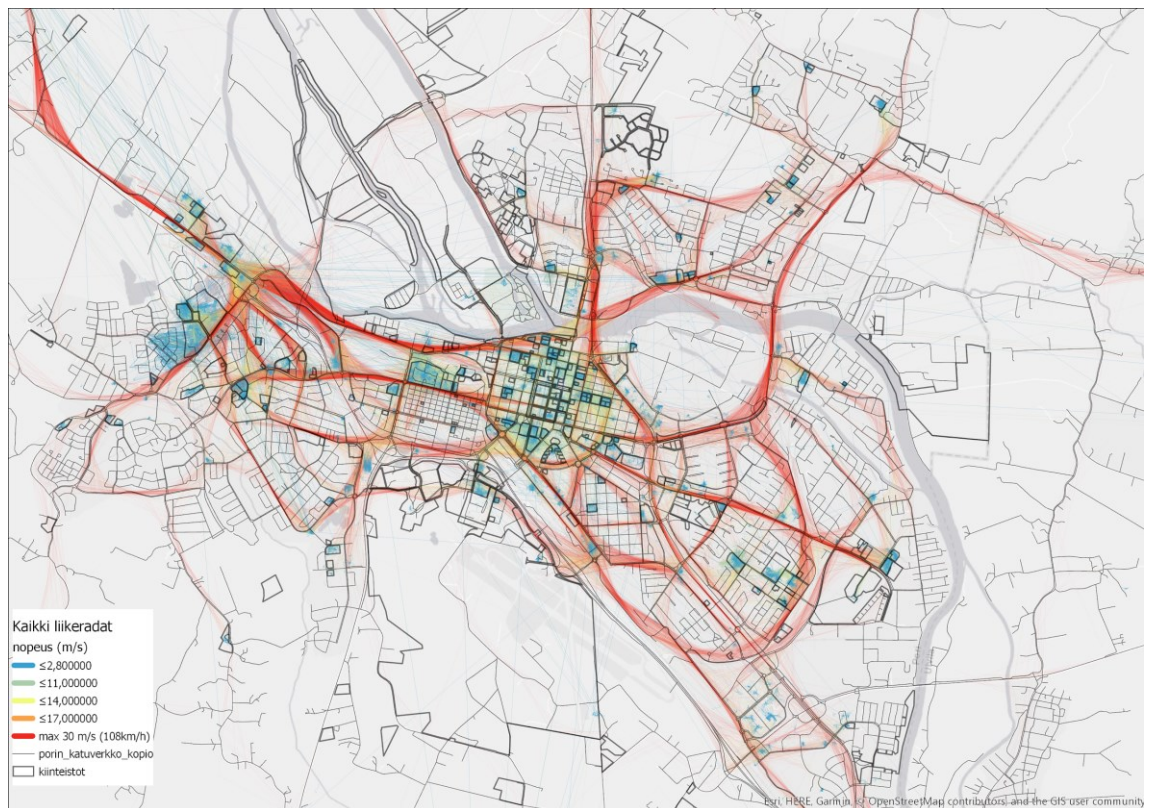
Tuloksista (Kuva 28 ja liite 4) voidaan nähdä luokittelun tulokset, missä liikeradan segmentit ovat visualisoitu nopeuden mukaan. Kun liikeratojen päälle laitetaan kiinteistö-



ja tieverkkoaineisto, nähdään selkeä korrelaatio nopeuden ja toiminnallisuuden välillä, jota voidaan tarkastella kartalla (Kuva 29).

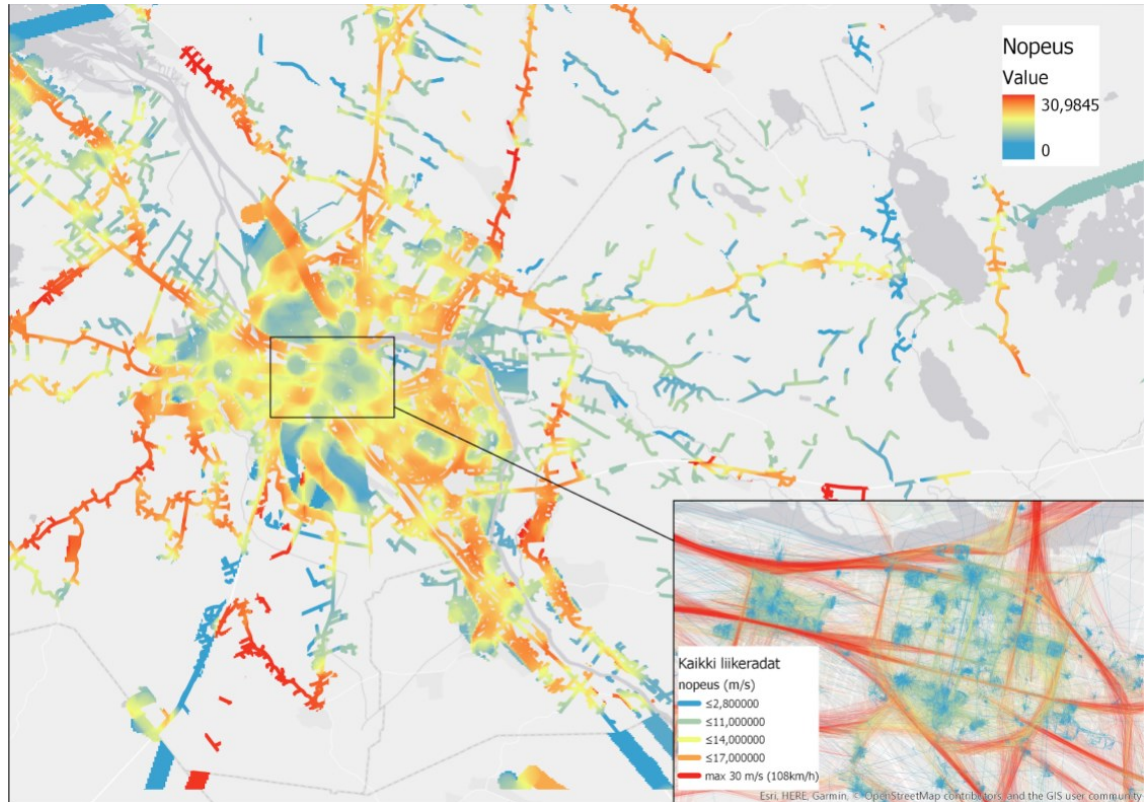


*Kuva 28: Kaikki liikeradat luokiteltu nopeuden mukaan.*



*Kuva 29: Kaikki liikeradat kiinteistöjen ja tieverkon kanssa. Siniset alueet ovat pääosin kiinteistöjen sisällä ja punaiset ovat tieverkon läheisyydessä.*

Liikeradat on vektoriaineisto ja luokittelun symboliikka vaikuttaa kartan tulkintaan, sillä symboliikassa pitää valita, missä järjestyksessä viivat esitetään. Kartoissa (Kuva 28 ja Kuva 29) päällimmäisenä esitetään hitaimmat liikeratasegmentit eli kaikki, jotka ovat alle 2.8 m/s. Tämä voimistaa visualisointia alueilla, joissa on hitaita nopeuksia. Kartasta on vaikea tulkita liikeratojen määrää eri luokissa suhteessa toisiinsa. Symbolointijärjestyksen ongelma poistuu, jos alueelle annetaan absoluuttinen keskiarvo kaikista nopeuksista niin, että myös alempana visualisoidut liikeradat ovat laskuissa mukana. Tämä saatiin aikaiseksi muuttamalla aineisto rastereiksi, joka kertoo solun alueen keskinopeuden. Kartassa (Kuva 30) on rasteriaineisto liukuarvoilla esitettynä samalla väriskaalalla kuin vektoriaineisto.

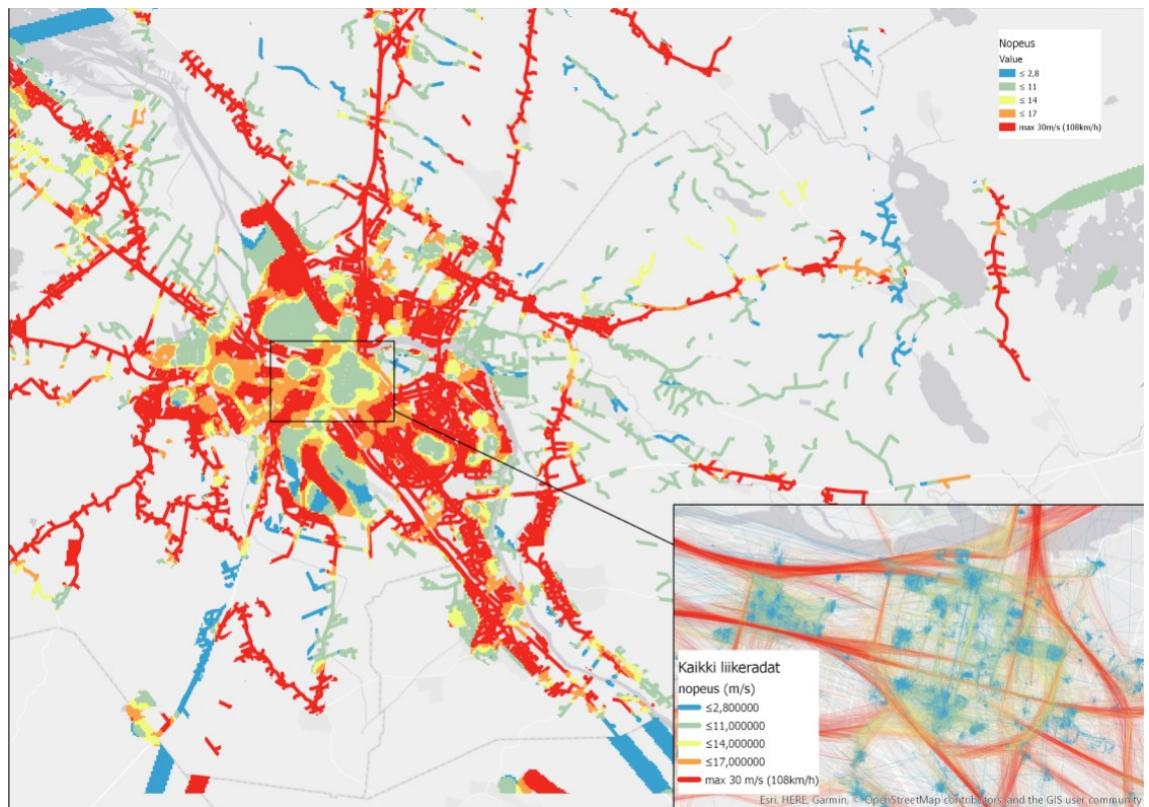


*Kuva 30: Nopeuden keskiarvo luokittelemattomana rasterina*

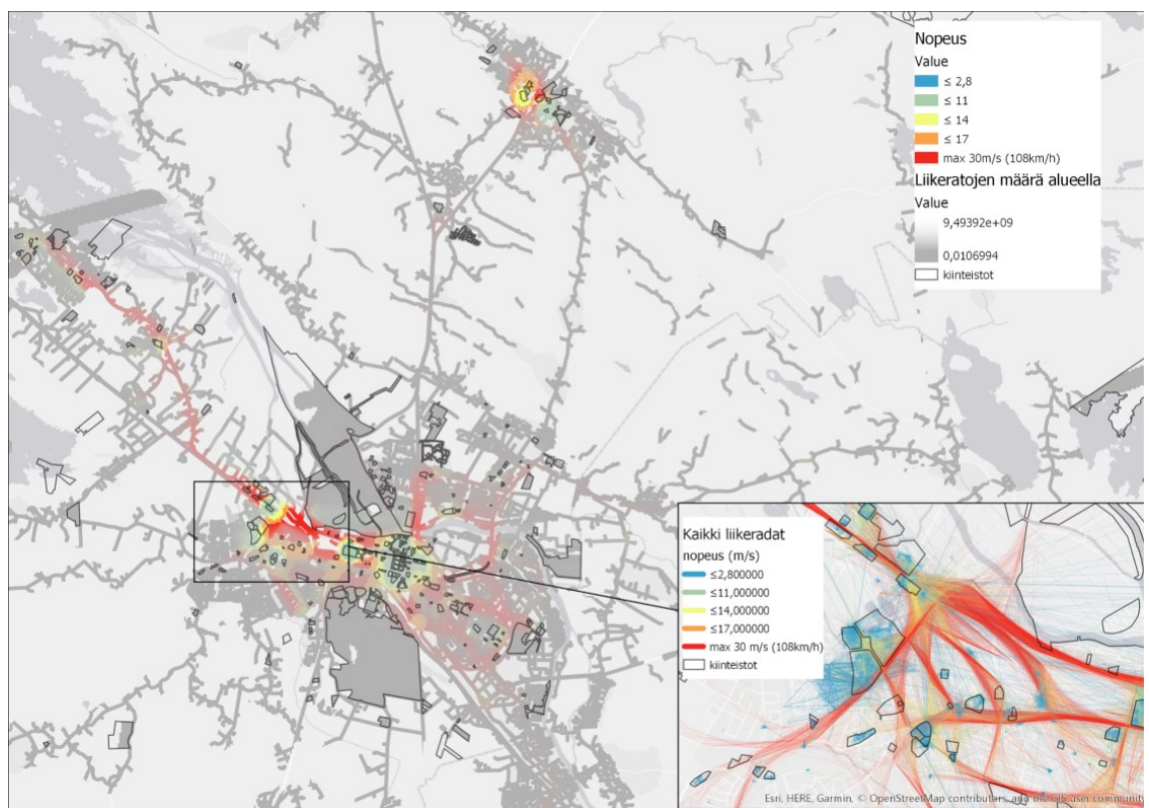
Kartassa (Kuva 31) rasteri on luokiteltu samoihin raja-arvoihin kuin vektoriaineisto: nopeuden mukaan viiteen luokkaan. Tässä esitystavassa ei enää erotu yksittäiset rakennukset tai tiet keskusta-alueelta. Rasteri on rajattu tieverkon ja kiinteistön mukaan.

Rasterin toinen vaikutus on, että liikeratojen määrää ei voi enää nähdä. Tämä ratkaistiin laskemalla toinen rasteri, joka summaa liikeratojen pituuksien summan solualueelta. Mitä enemmän liikeratoja löytyy alueelta, sitä voimakkaammin värit tulevat harmaan läpi. Kyseinen esitystapa on seuraavissa kartoissa (Kuva 32 ja Kuva 33).

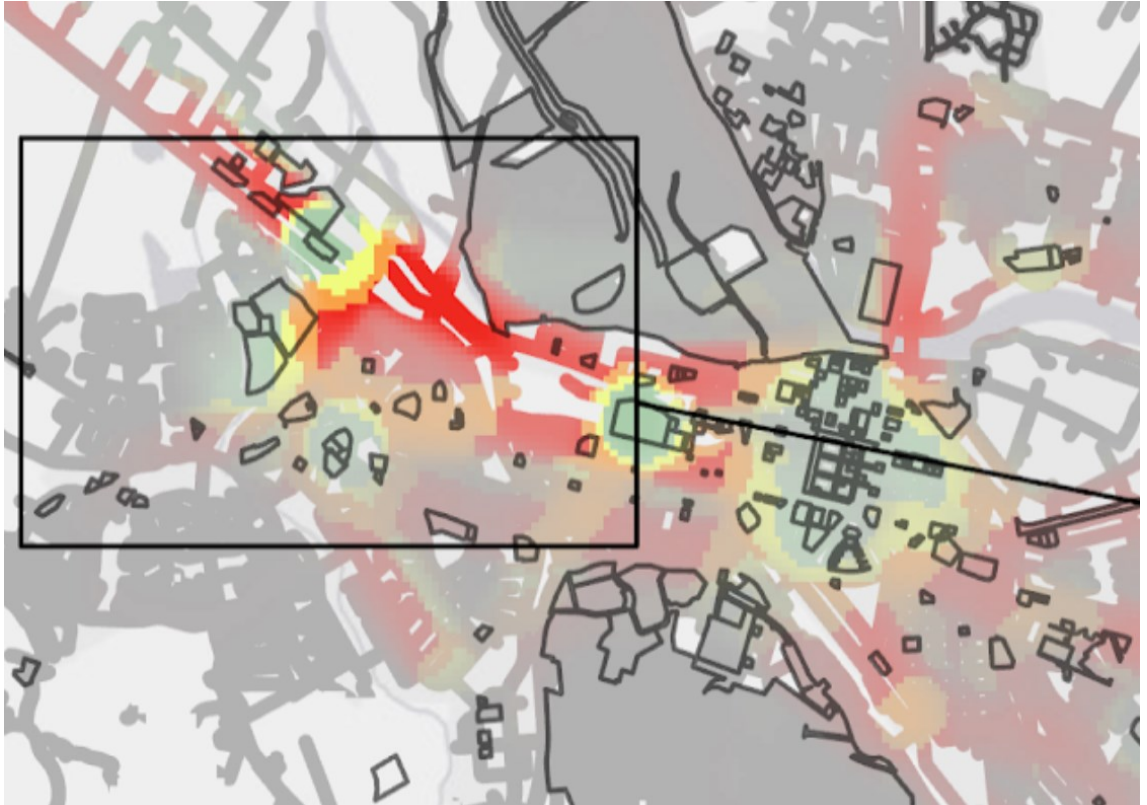




Kuva 31: Nopeuden keskiarvo rasterina ja uudelleen luokiteltuna

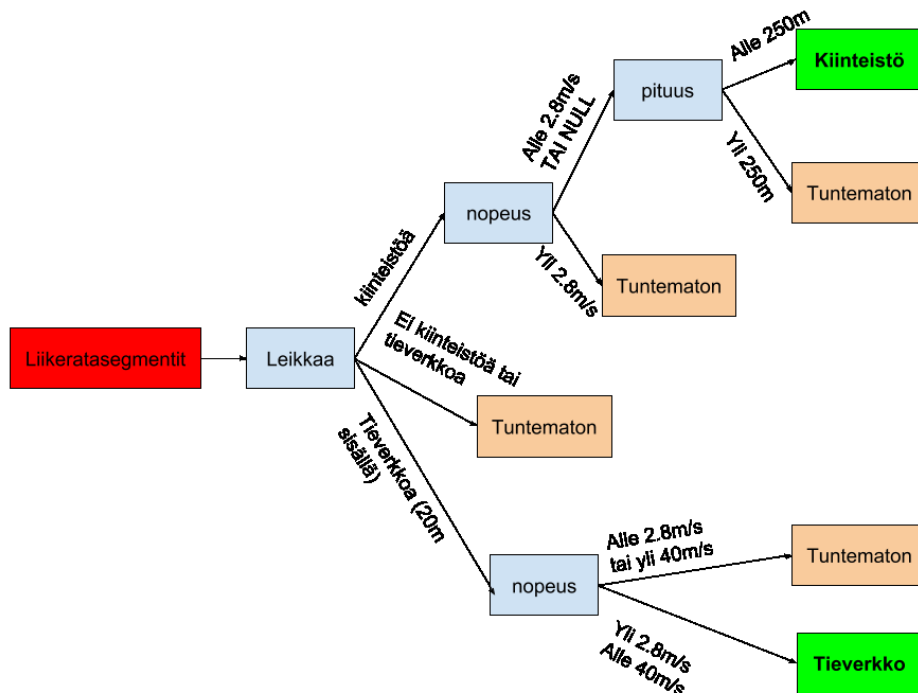


Kuva 32: Kaksi rasteria päällekkäin, missä toinen kertoo nopeuden ja toinen liikeratojen määrän alueella



Kuva 33: Lähikuva rasterista, joka on Kuva 32.

Nopeuden lisäksi otettiin huomioon myös liikeratojen segmenttien sijainti, leikkaavatko ne kiinteistöjä tai tieverkkoa 20 metrin etäisyydeltä. Luokittelumenetelmäksi valittiin päätöspuu ja luokittelusäännöt. Lopullisessa luokittelussa on 3 luokkaa: Kiinteistö, Tieverkko ja Tuntematon. Päätöspuu, joka luokittelee liikeratasegmentit, on alla olevassa kuvassa (Kuva 34).



Kuva 34: Liikeratojen luokittelun päätöspuu



Päätöspuun pohjalta muodostettiin luokittelusäännöt. Luokittelusäännöt ovat IF-THEN muodossa ja ovat seuraavalla tavalla:

**Muuttujat:**

Liikeradat = Kaikki liikeradat, leikattu segmentteihin nopeuden mukaan

Kiinteistöt = Porin kaupungin kiinteistöt

Nopeus = m/s, laskettu valmiiksi jokaiselle liikeradan segmentille

Pituus = Liikeradan segmentin pituus metreinä.

**Kiinteistö:**

If Liikeradat intersect Kiinteistöt AND (Nopeus  $\leq$  2.8 OR Nopeus is NULL) AND Pituus  $<$  250:

Then "Kiinteistö"

**Tieverkko:**

If Liikeradat intersect Tieverkko within 20 meters AND Nopeus  $>$  2.8 AND Nopeus  $<$  40:

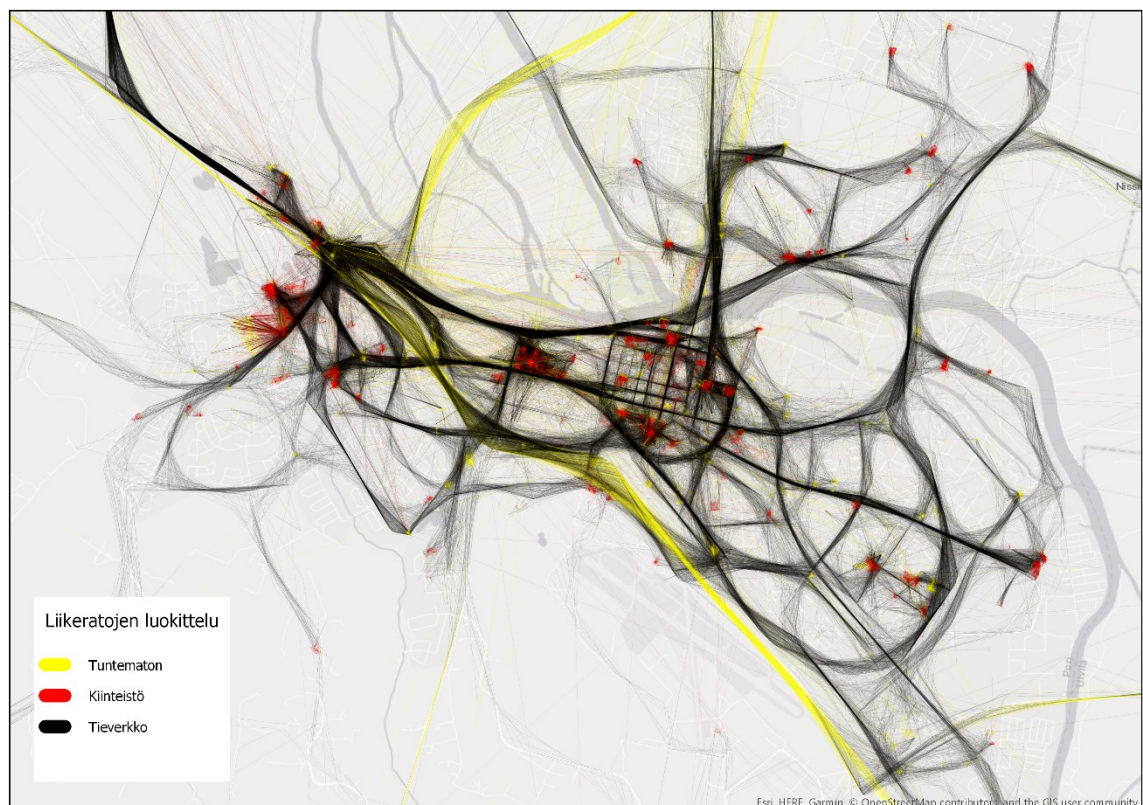
Then "Tieverkko"

**Tuntematon:**

If Liikerata is not Kiinteistö AND Liikerata is not Tieverkko:

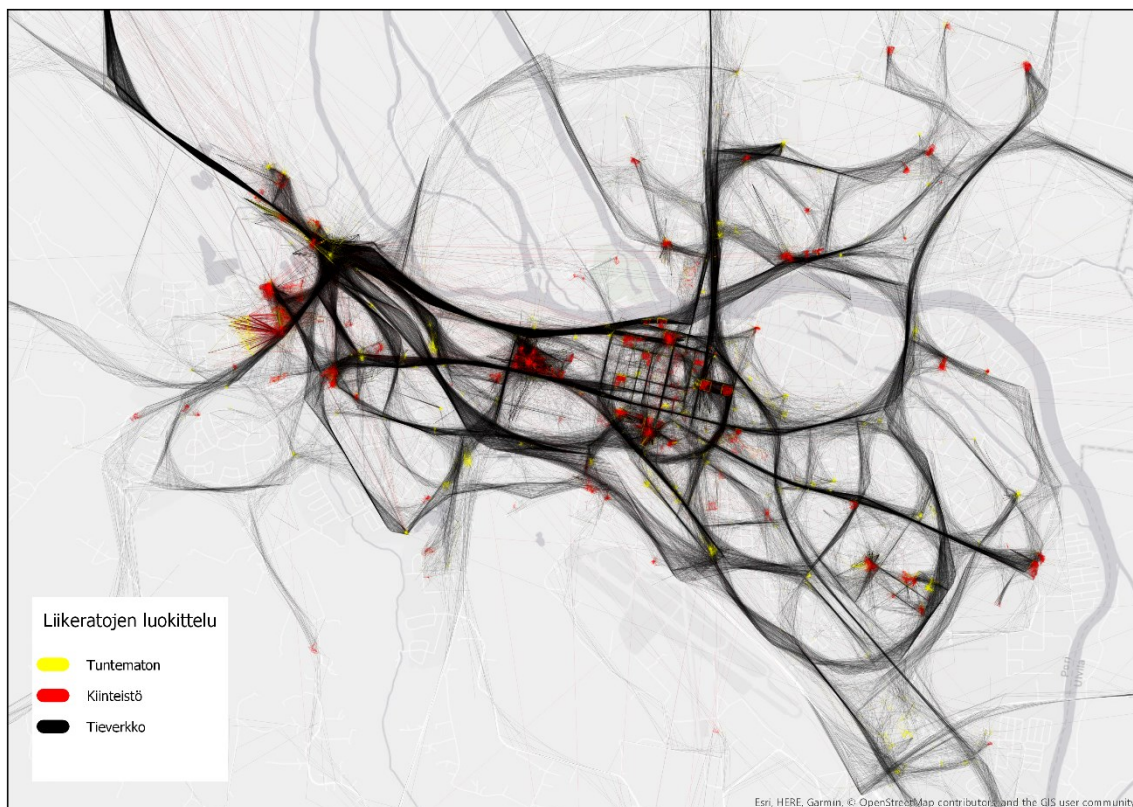
Then "Tuntematon"

Luokittelupuun ja luokittelusääntöjen avulla luokitellut liikeradan segmentit ovat visualisoitu kolmella värillä (Kuva 35). Kuvassa näkyy kaikkien liikeratojen luokittelutulokset. Tuntemattomaan luokkaan jäi suuri määrä liikeratoja, jotka visuaalisesti tulkittuna kuuluvat tieverkkoiluokkaan, mutta niiden nopeus on yli 40 m/s tai ne ovat liian kaukana tieverkosta.

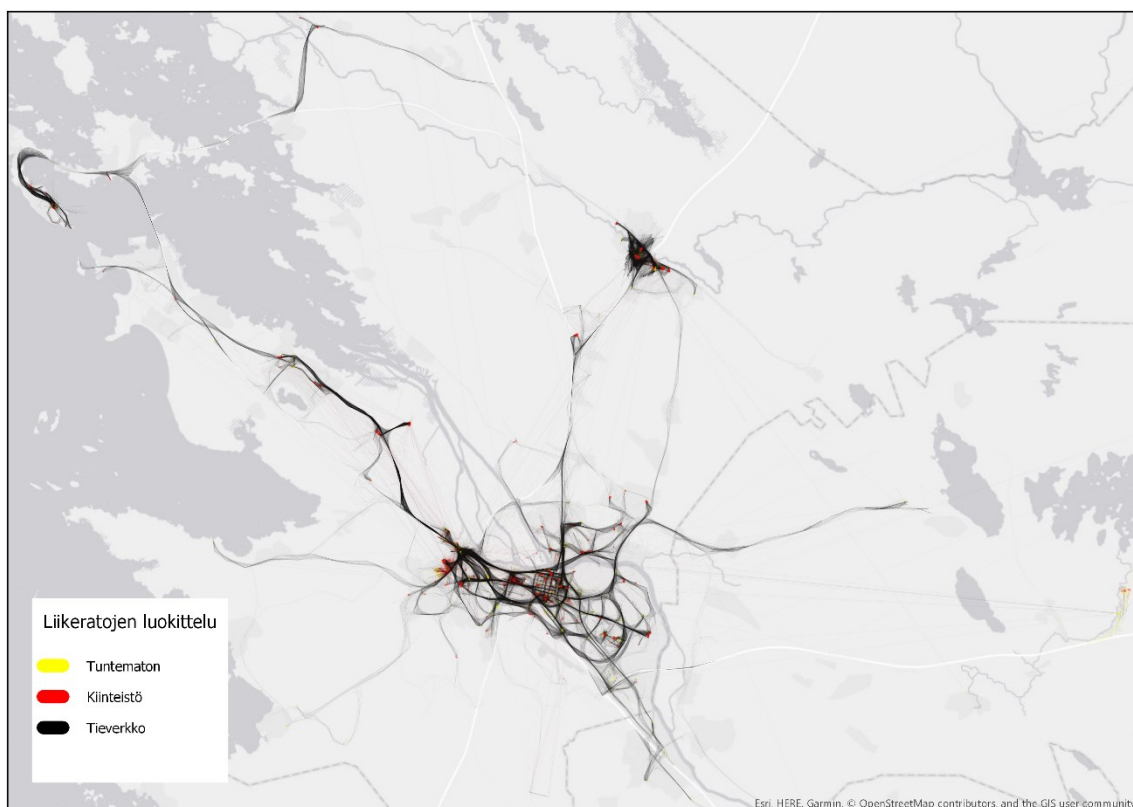


Kuva 35: Liikeratojen luokittelu

Seuraavaksi suodatetaan häiriöpisteet pois, eli kaikki yli 40 m/s nopeudella olevat segmentit. Tulokset ovat seuraavissa kartoissa (Kuva 36 ja Kuva 37).

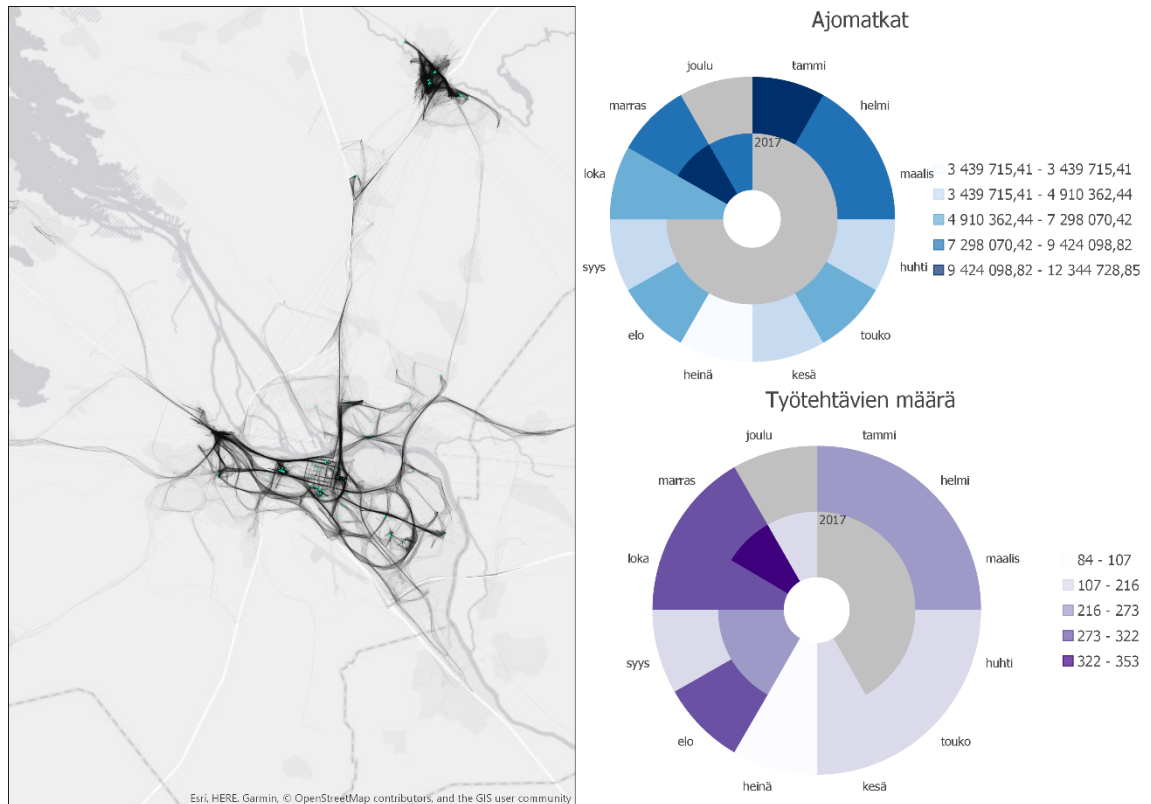


*Kuva 36: Kaikki yli 40 m/s suodatettu pois, kartta keskustan alueelta.*



*Kuva 37: Yli 40 m/s suodatettu pois, kartta koko aineiston alueelta.*

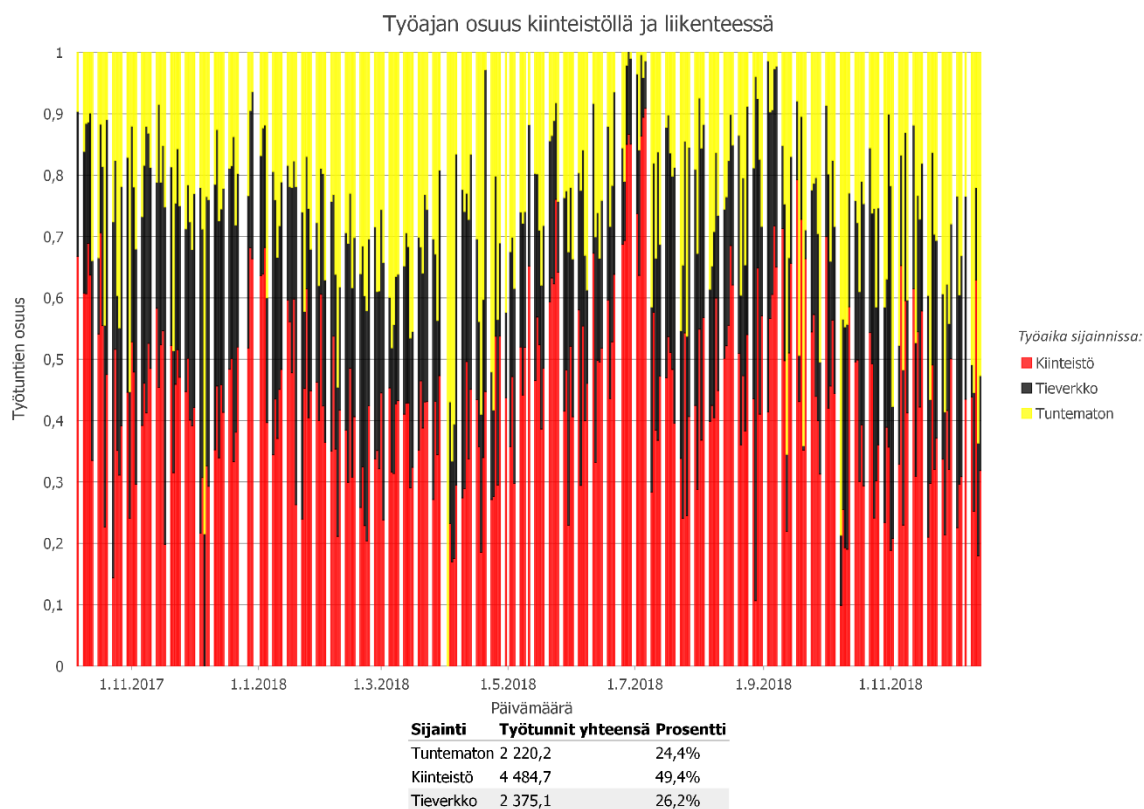
Luokittelun avulla voidaan mitata ajomatkaa ja verrata tuloksia halutulta ajanjaksolta. Kuva 38 esittää ajomatkaa kuukausittain kellomaisessa kaaviossa. Kaavion visualisointimenetelmä on luonnollisen jakauman -algoritmin mukainen ja on summa liikeratasegmenteistä, jotka ovat luokiteltu tieverkkoon. Luvut ovat metreissä.



Kuva 38: Ajomatkat ja työtehtävät kuukausittain

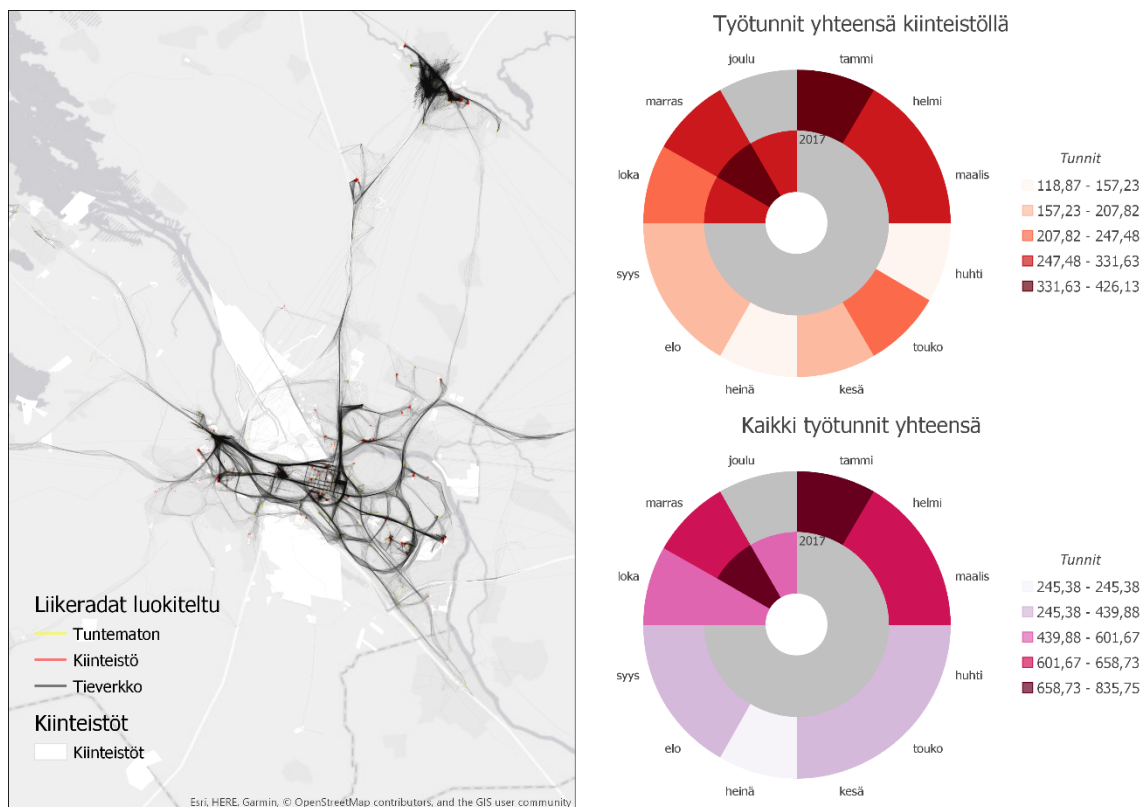
Luokittelun avulla voidaan myös määritellä työtuntien määrä, niin ajon aikana kuin kiinteistöllä halutulla ajanjaksolla. Työajan luokittelu on esitetty kaaviossa (Kuva 39). Kaaviossa on myös taulukko, missä on esitetty työtunnit yhteensä sekä niiden prosentuaalinen osuus. Tulokset osoittavat, että noin puolet työajasta on huollettavilla kiinteistöllä ja 26 % matkustamisessa. Loput työajasta on seurattavien työtehtävien ulkopuolista toimintaa sekä luokittelematonta aineistoa merkitty tuntemattomaksi. Lopullisessa luokittelussa tuntemattomaksi jäi enää pieniä rajattuja alueita, jotka viittaavat pääasiassa tuntemattomiin kiinteistöihin. Työntekijöiden tehtäviin kuuluu myös muita kiinteistöhuoltotoimia, jotka eivät ole tämän projektin työtehtäväkuvauksissa.





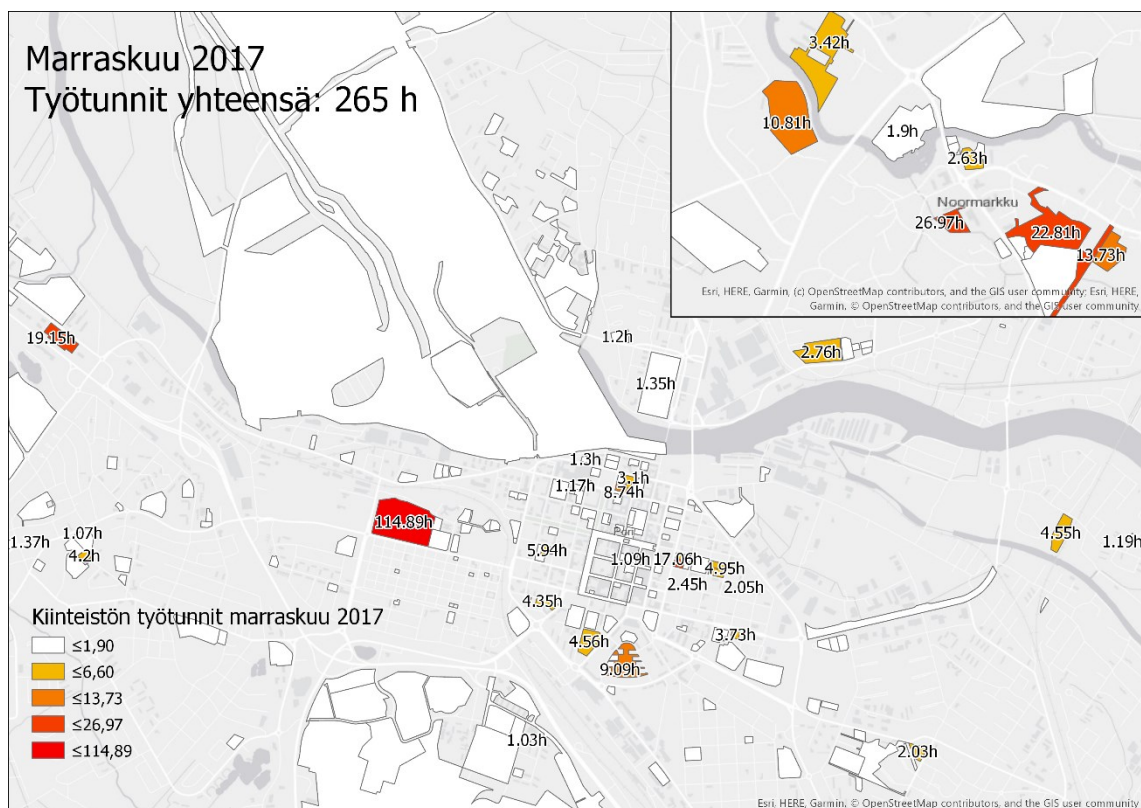
Kuva 39: Työaika luokittelulla jaettuna

Työajanmuutosta ajansuhteen voidaan esittää myös kellomaisella kaaviolla, jota kuva 40 esittää. Kuvassa on eroteltu kiinteistöllä oleva työaika muusta ajasta luokittelun avulla.

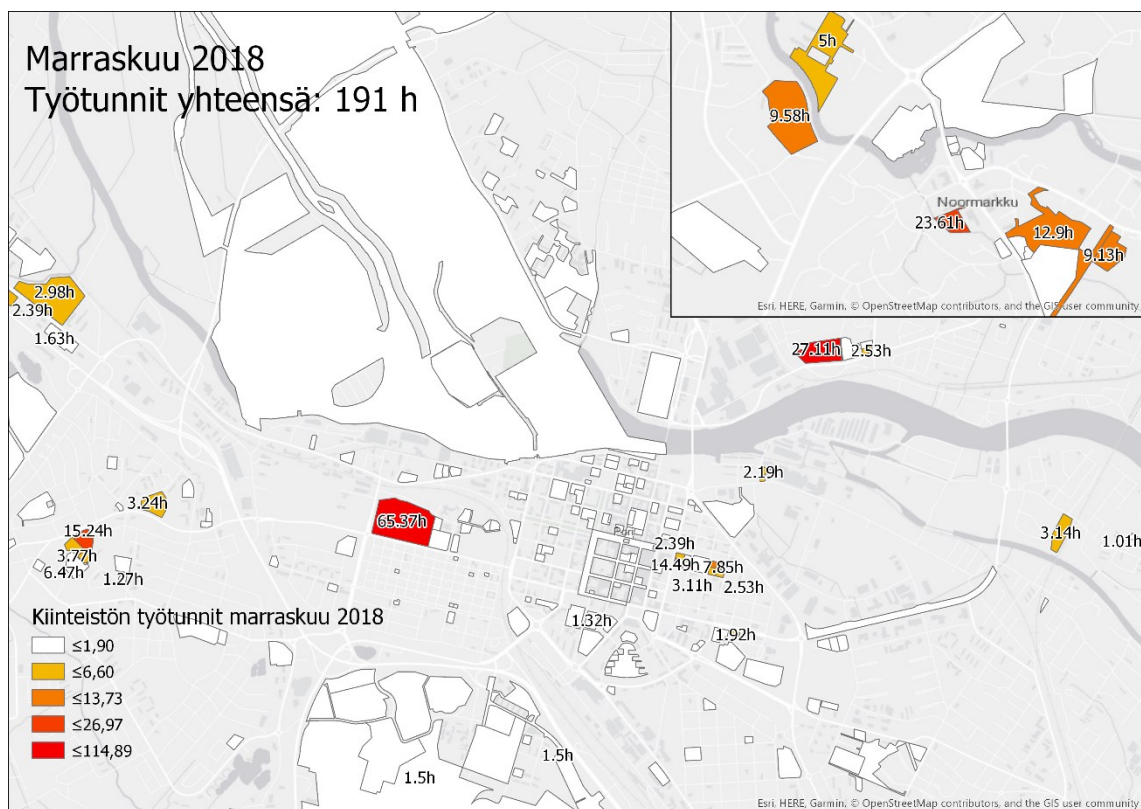


Kuva 40: Työaika kiinteistöllä ja kaikki yhteensä kuukausittain.

Työtunnit kuukausi- ja kiinteistötasolla tehtiin summaamalla kiinteistöiksi luokitellut liikeratasegmentit kiinteistörajojen mukaan (Kuva 41 ja Kuva 42).



Kuva 41: Marraskuu 2017 työtunnit kiinteistöittäin

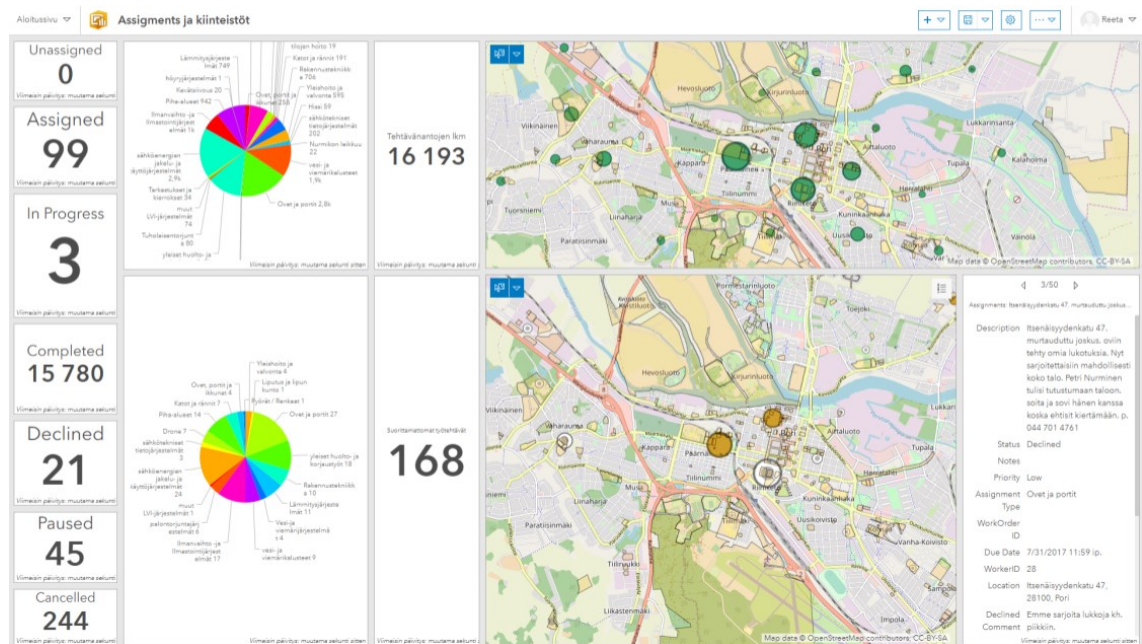


Kuva 42: Marraskuu 2018 työtunnit kiinteistöittäin

### 4.2.3 Visuaalisten menetelmien tulokset

Reaaliaikaisen aineiston tulkitsemiseen toimii hyvin moniosainen ja interaktiivinen sovellus, missä käyttäjä itse voi päättää, mitä aluetta tarkastelee. Alla olevissa kuvissa on kuvakaappauksia (Kuva 43, Kuva 44 ja Kuva 45) Operation Dashboard -sovelluksesta, missä on työtehtävät ja kiinteistöt kartalla. Työtehtävien tehtävätyyppi on piirakkakaavioissa ja numeroarvoina löytyy työtehtävien tila.

Sovellus antaa yhteenvedon tehtävänannoista suhteessa kiinteistöihin. Sovelluksessa on kaksi karttaa. Ylemmässä kartassa on kaikki tehtävänannot koko aineiston keräysajalta ja alemmassa kartassa on nykyiset avoimet työtehtävät. Tehtävänannot ovat visualisoitu klusteroinnilla ja lähemmässä tarkastelussa tulee kaikki pisteet näkyviin. Sovelluksessa on karttojen lisäksi informaatioikkunat, missä piirakkakaaviot esittävät tehtävätyypin ja sen osuuden muihin nähden. Numeroarvioikkunat esittävät tehtävänantojen määrää tilan mukaan. Sovellus on myös interaktiivinen sillä, jos valitsee kartalta kiinteistön, informaatioikkunat päivittyvät ja näyttävät kyseisen kiinteistön alueella olevat tehtävänannot. Lisäksi, jos valitsee piirakkakaaviosta yhden sektorin, se suodattaa kartalta sen tehtävätyypin mukaiset pisteet ja esittää vain ne kartalla. Sovelluksen käyttötarkoitus on seurata tehtävätyypien ja tilojen kokonaismäärää suhteessa kiinteistöihin. Tärkeimmät luvut on esitetty numeroiden avulla ja työnohjaukseen saa reaaliaikaisen tiedon kokonaistilanteesta.



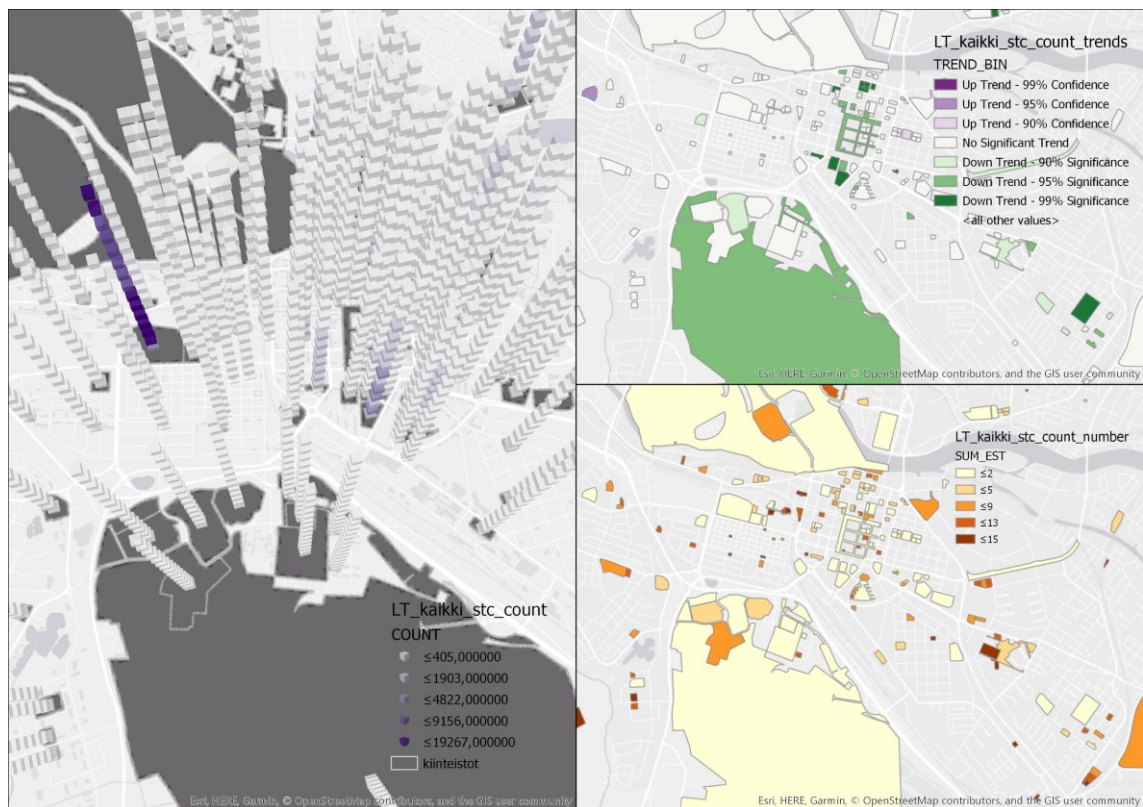
Kuva 43: Operation Dashboard, sovellus työtehtävien seurantaan, aloitusnäkymä.





noin 19 200 pistettä yhden kuukauden aikana, mikä näkyy kuvassa violettina laatikoina. Yhdellä kiinteistöllä oli selvästi tasaisesti eniten käyntejä ja se näkyy kuvassa vasemmalla violettina pylväänä. Aika-avaruuskuutiot toimivat kolmiulotteisessa tarkastelussa parhaiten dynaamisena aineistona ja siinä voidaan rajata haluttu ajanjakso tai määrä näkyviin. Staattisissa kartoissa toimii parhaiten kaksiulotteiset kartat, jotka kertovat aineistosta trendejä ajansuhteen.

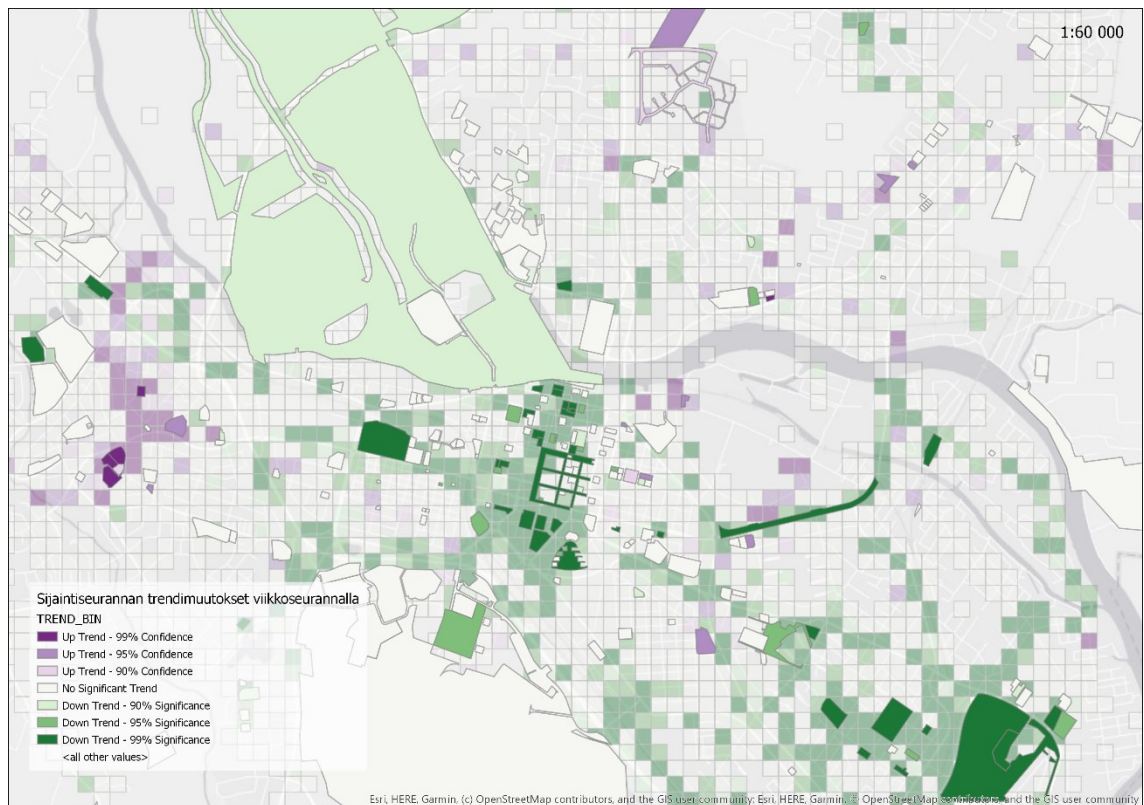
Kuvassa (Kuva 46) oikealla on kaksi analyysiä aika-avaruuskuutiosta, joista ylempi kertoo, onko kiinteistöllä vähentynyt tai kasvanut käynnit suhteessa aiempiin käynteihin. Alempi kuva kertoo, kuinka monen kuution arvo on jouduttu arvioimaan, sillä pisteitä ei ole ollut kyseisen alueen sisällä kyseisellä ajanjaksolla.



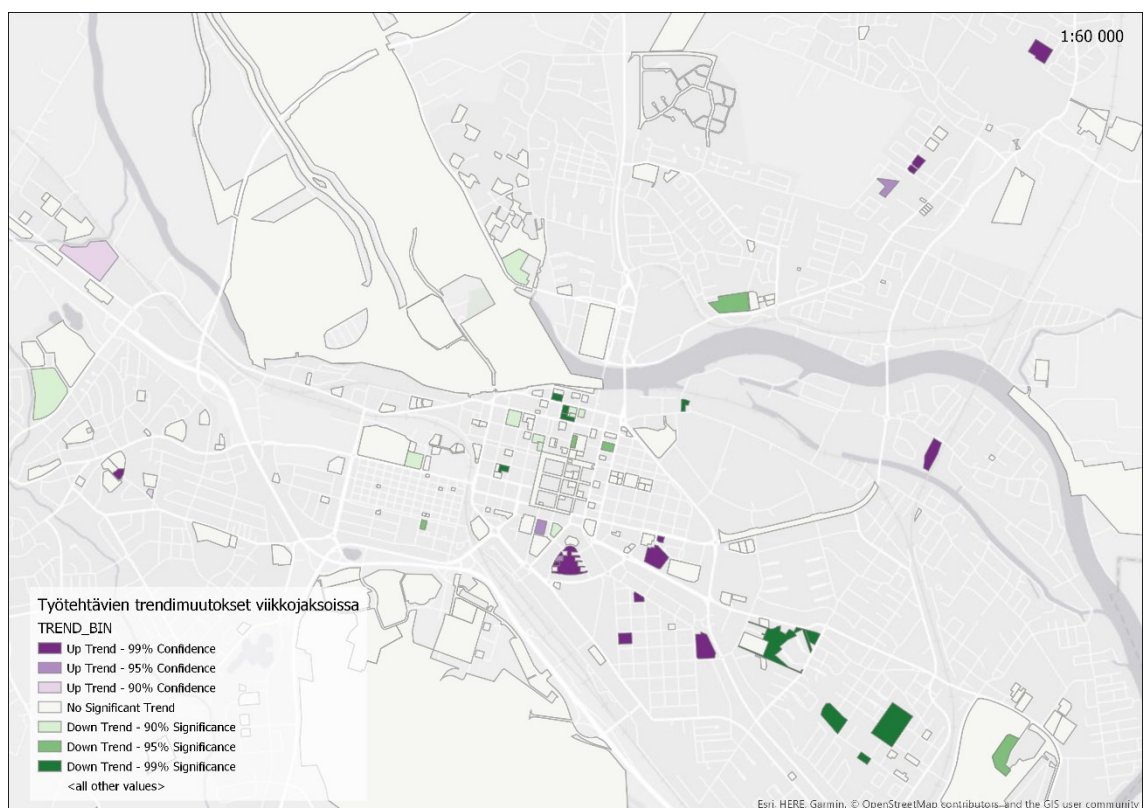
Kuva 46 Esimerkki aika-avaruuskuutioiden visualisoinnista.

Näillä analyyseillä pystymme löytämään kiinteistökohtaisesti trendejä aineistosta kiinteistöihin liittyen. Esimerkiksi, jos kiinteistöllä on “Up trend”, niin kiinteistön huoltokustannukset ovat nousseet ja jos kiinteistöllä on “Down trend”, kiinteistön huoltokustannukset ovat laskeneet. Tämän saman ilmiön pystyy vahvistamaan ja tarkentamaan analysoimalla annettujen työtehtävien trendejä. Seuraavat kartat (Kuva 47 ja Kuva 48) esittävät muutostrendiä sijaintiseurannassa ja työtehtävien aloittamisessa samalta alueelta. Molemmissa on analysoitu kiinteistökohtaisilla aikasarjoilla muutostrendiä. Sijaintiseurannassa on lisäksi analysoitu 250 metrin välein muutostrendiä aikasarjoissa. Näitä kahta karttaa vertailemalla, huomataan, että muutostrendi ei korreloi niin vahvasti kuin odotettu tulos. Työtehtävillä on useammin nouseva muutostrendi verrattuna sijainnin seurannan muutostrendeihin, jotka ovat yleisesti laskevia melkein joka alueella. Muista alueista tehdyt tulokset löytyvät liitteestä 5.





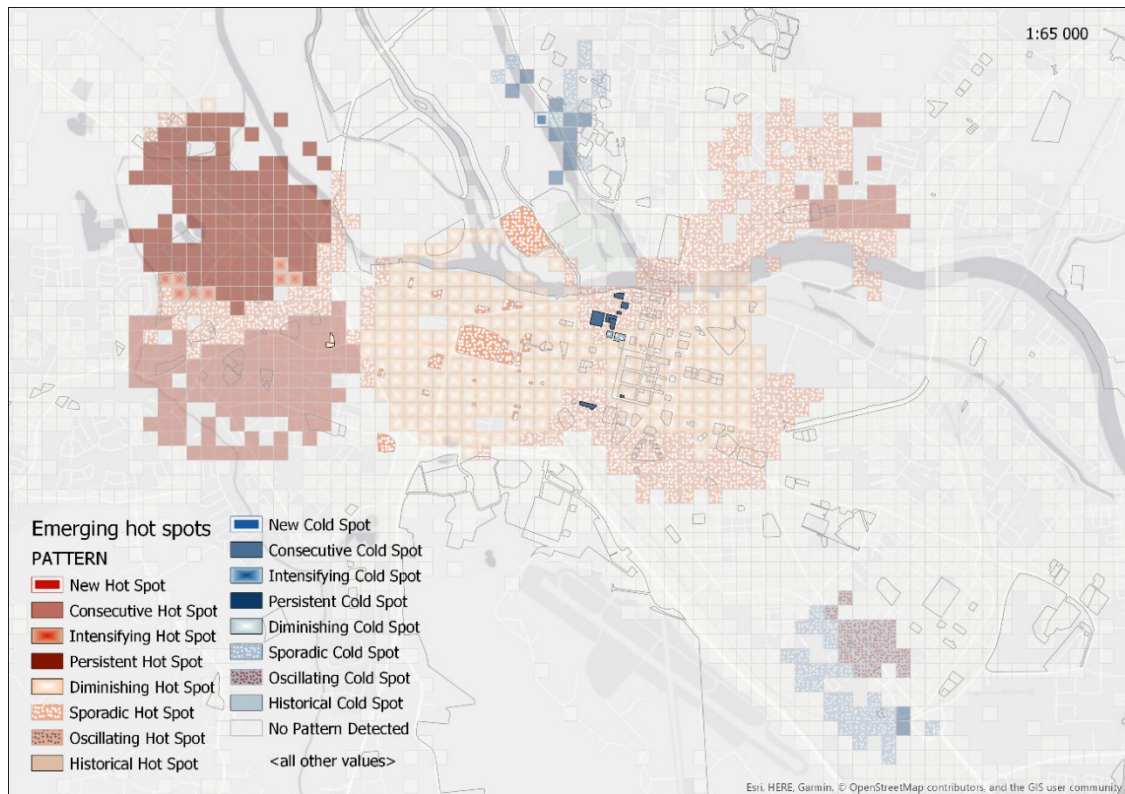
Kuva 47: Sijaintiseurannan muutostrendit 250 m ruuduissa ja kiinteistökohtaisesti viikkotasolla laskettuna.



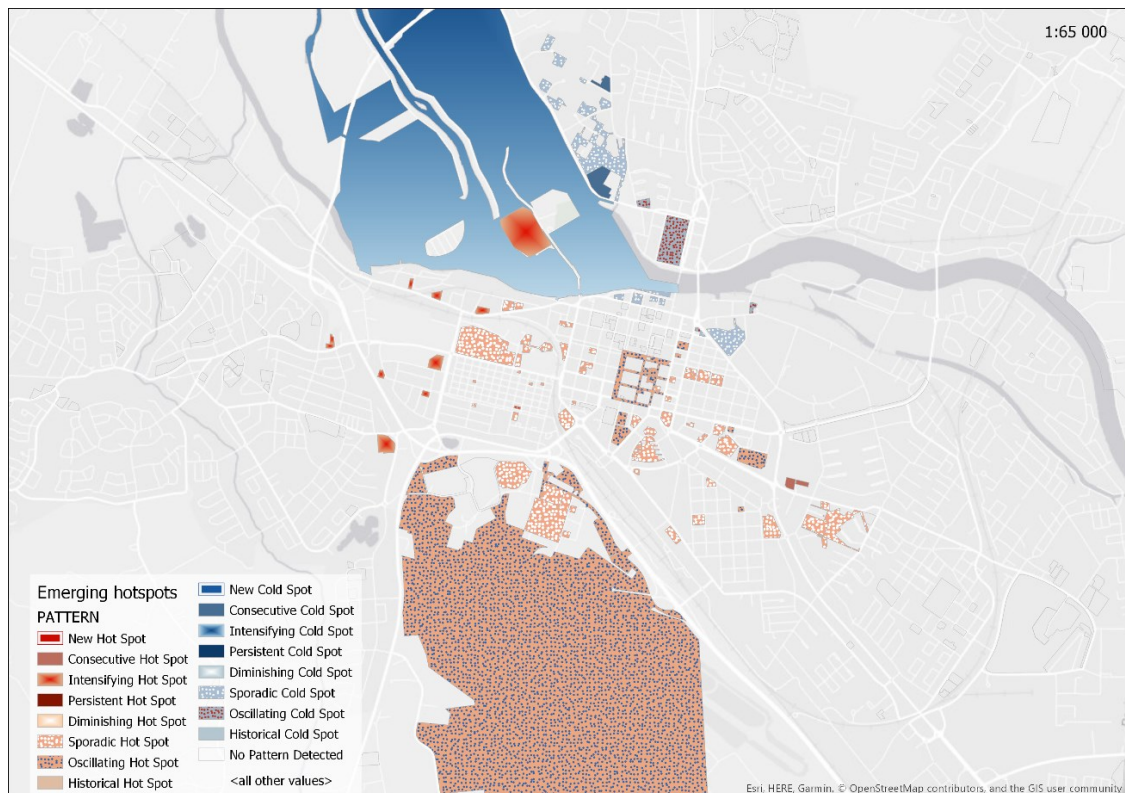
Kuva 48: Työtehtävien muutostrendit viikkotasolla laskettuna.

Muutostrendin lisäksi aika-avaruuskuutioista voidaan analysoida kuuma- ja kylmäpisteitä. Kartat (Kuva 49 ja Kuva 50) esittävät analyysin tuloksia keskustan alueelta seurattavien sijaintien ja tehtävänantojen osalta. Kartan yleisilme on punertava, mikä

viittaa kuumapisteisiin. Analyysi erottelee erilaiset kuuma- ja kylmäpisteet toisistaan ja taulukko 4 esittelee tarkemman selityksen erilaisten kuuma- ja kylmäpistealueiden symboleille.




















Kuva 49: Kuuma- ja kylmäpisteanalyysi



Kuva 50: Kuuma- ja kylmäpisteanalyysitulokset tehtävänannoille

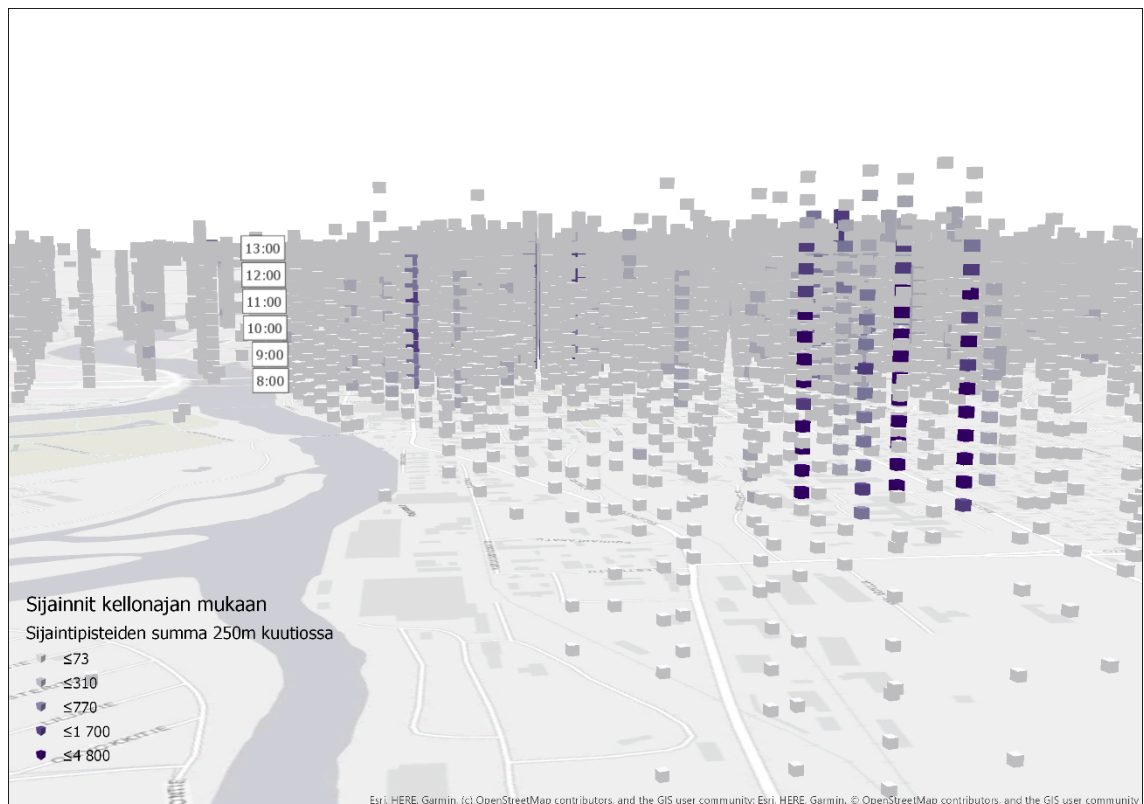


*Taulukko 4: Kuuma- ja kylmäpisteanalyysin tuloksien selitys*

Kuvio	Kuvion nimi	Suomennettu selitys
	No Pattern Detected	Ei ole kuuma- tai kylmäpiste. Ei tilastollisesti merkittävää muutosta tai klusterointia.
	New Hot Spot	Tilastollisesti merkittävä kuumapiste vain viimeisessä aikajaksossa, muutoin ei ole ollut merkittävä sijainti.
	Consecutive Hot Spot	Sijainnissa on ollut yhtäjaksoisesti kuumapisteitä uusimpien ajanjaksojen ajan. Sijainti ei ole muuten ollut tilastollisesti kuumapiste ja alle 90% kuutioista on tilastollisesti merkittäviä kuumapisteitä.
	Intensifying Hot Spot	Sijainti on ollut tilastollisesti merkittävä kuumapiste 90% ajanjaksoista, mukana lukien viimeinen ajanjakso. Lisäksi intensiivisyys on ollut nousussa.
	Persistent Hot Spot	Sijainti on ollut tilastollisesti merkittävä kuumapiste 90% ajanjaksoista ja intensiivisyys ei ole noussut tai laskenut merkittävästi.
	Diminishing Hot Spot	Sijainti on ollut tilastollisesti merkittävä kuumapiste 90% ajanjaksoista, mukana lukien viimeinen ajanjakso. Lisäksi intensiivisyys on ollut laskussa.
	Sporadic Hot Spot	Sijainti on välillä kuumapiste ja välillä ei, mutta vähemmän kuin 90% ajanjaksoista kuumapiste. Lisäksi se ei ole koskaan ollut kylmäpiste.
	Oscillating Hot Spot	Sijainti on viimeksi ollut kuuma piste, mutta historiallisesti myös kylmä piste. Se on ollut vähemmän kuin 90% ajasta kuuma piste.
	Historical Hot Spot	Uusin ajanjakso ei ole kuuma piste, mutta vähintään 90% ajanjaksoista se on ollut kuuma piste.
	New Cold Spot	Tilastollisesti merkittävä kylmäpiste vain viimeisessä aikajaksossa, muutoin ei ole ollut merkittävä sijainti.
	Consecutive Cold Spot	Sijainnissa on ollut yhtäjaksoisesti kylmiä pisteitä uusimpien ajanjaksojen ajan. Sijainti ei ole muuten ollut tilastollisesti kylmäpiste ja alle 90% kuutioista on tilastollisesti merkittäviä kylmäpisteitä.
	Intensifying Cold Spot	Sijainti on ollut tilastollisesti merkittävä kylmäpiste 90% ajanjaksoista, mukana lukien viimeinen ajanjakso. Lisäksi intensiivisyys on ollut nousussa.
	Persistent Cold Spot	Sijainti on ollut tilastollisesti merkittävä kylmäpiste 90% ajanjaksoista ja intensiivisyys ei ole noussut tai laskenut merkittävästi.
	Diminishing Cold Spot	Sijainti on ollut tilastollisesti merkittävä kylmäpiste 90% ajanjaksoista, mukana lukien viimeinen ajanjakso. Lisäksi intensiivisyys on ollut laskussa.
	Sporadic Cold Spot	Sijainti on välillä kylmäpiste ja välillä ei, mutta vähemmän kuin 90% ajanjaksoista kylmäpiste. Lisäksi se ei ole koskaan ollut kuumapiste.
	Oscillating Cold Spot	Sijainti on viimeksi ollut kylmäpiste, mutta historiallisesti myös kuumapiste. Se on ollut vähemmän kuin 90% ajasta kylmäpiste.
	Historical Cold Spot	Uusin ajanjakso ei ole kylmäpiste, mutta vähintään 90% ajanjaksoista se on ollut kylmäpiste.

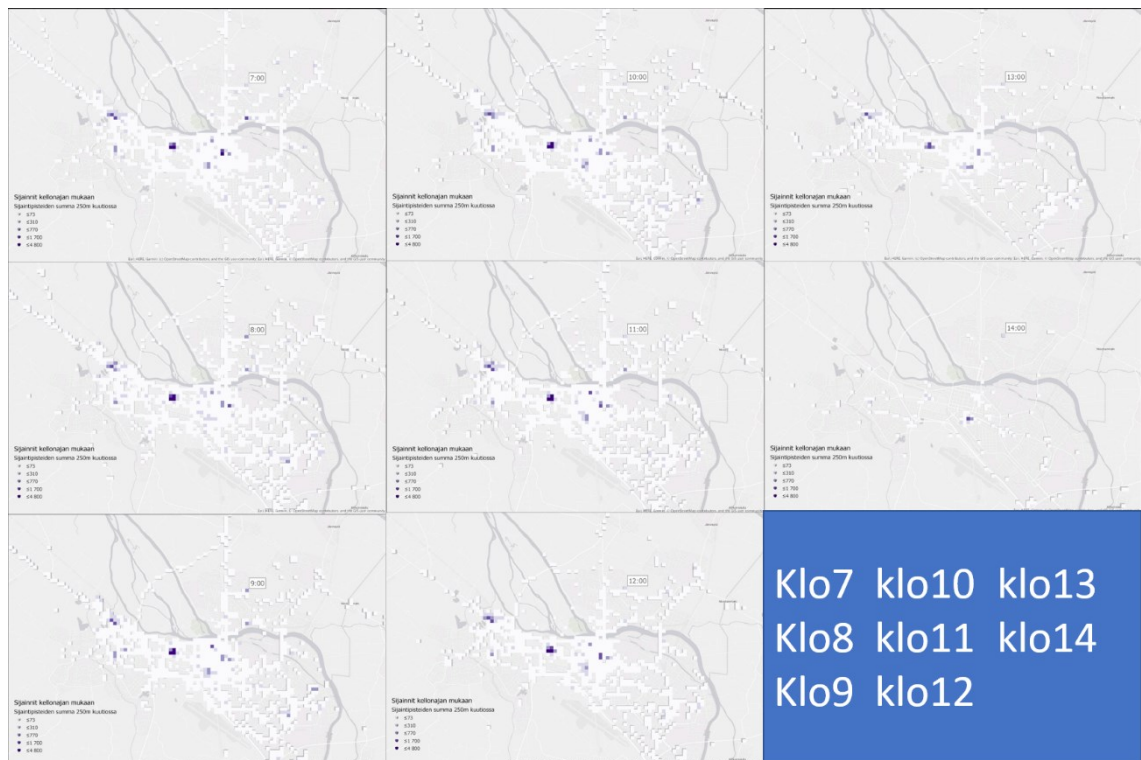
Kartat (Kuva 49 ja Kuva 50) tulokset viittaavat taulukon selityksien mukaan siihen, että keskustan alueella on yleisesti ollut enemmän kuumapisteitä, mutta useassa sijainnissa viimeisin ajanjakso ei ole ollut kuumapiste. Aineiston viimeinen ajanjakso on 23.-30.11.2018. Lisäksi sijaintiseurannassa keskustan alueella on muutama selkeä kylmäpiste, mikä viittaa siihen, että kylmäpistealueella olevat kiinteistöt ovat tilastollisesti merkittävästi vähiten vierailtuja kohteita verrattuna ympärillä oleviin kiinteistöihin. Nämä samat kiinteistöt tehtävänantojen analyysissä eivät ole merkittäviä pisteitä, mutta tämä voi johtua siitä, että ympärillä ei ole selkeitä ja voimakkaita kylmä- tai kuumapisteitä, jotka tekisivät kyseisistä sijainneista tilastollisesti merkittäviä ympäristöönsä nähden. Analyysin parametreiksi annettiin paljon painoarvoa spatiaalisille ja temporaalisille naapureille, sillä analyysissä vertailtiin 24 spatiaaliseen naapuriin (2 km maksimisäteellä) ja 8 temporaaliseen naapuriin. Nämä parametrit voimistivat tuloksia, jotta saadaan selkeämpiä eroja alueiden välille. Muista alueista löytyy tulokset liitteistä 6.

Viikkotarkastelun lisäksi tarkasteltiin kellonajan vaikutusta muutostrendeihin sekä kuuma- ja kylmäpisteanalyysiin. Lähtöaineisto muutettuna aika-avaruuskuutioiksi on esitetty kolmiulotteisena kartalla (Kuva 51). Kolmiulotteinen esitystapa ei ole hyödyllinen staattisena kuvana, joten useat kellonajan hetket ovat esitettynä kaksiulotteisena (Kuva 52 ja liitteessä 7).



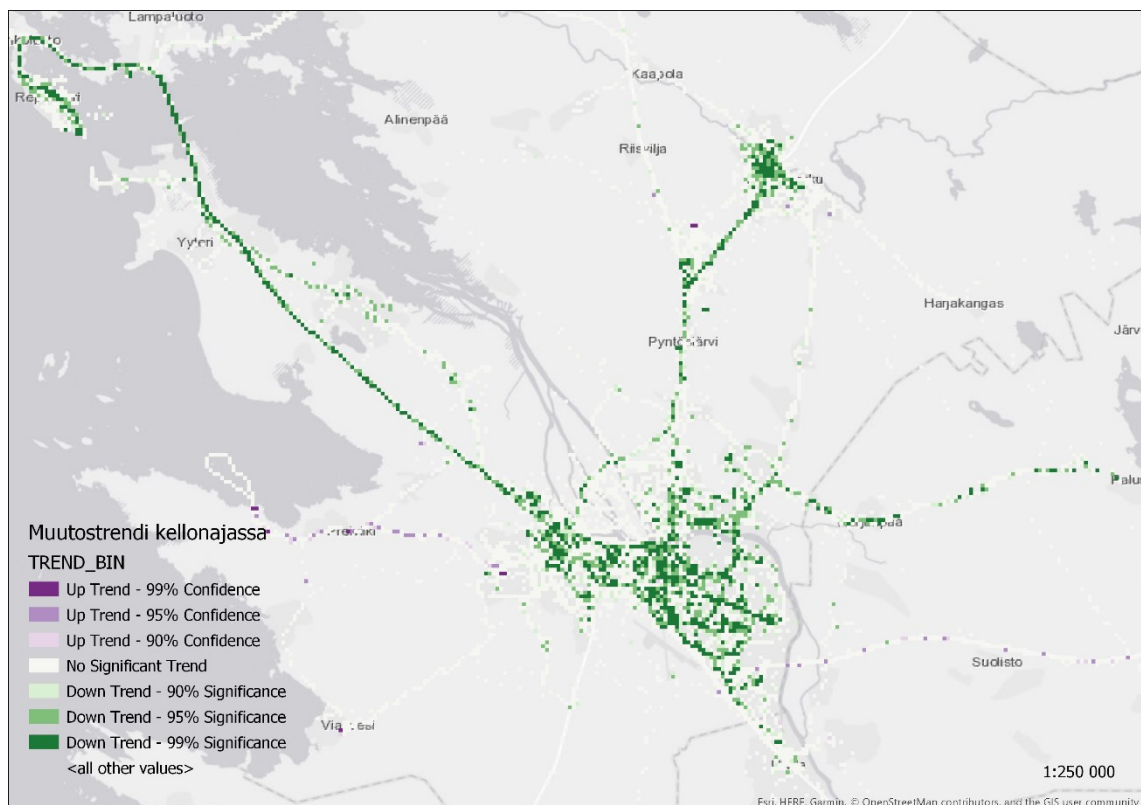
*Kuva 51: Sijaintiseuranta kellonajan mukaan 250 m aika-avaruuskuutioissa*

Kuva 52 ja liitteen 7 aikasarjaesityksessä kellonaikojen välillä ei ole suurta eroa, paitsi viimeisen tunnin osalta, missä kokonaismäärä pisteitä ja sijainteja on vähentynyt huomattavasti. Tulokset osoittavat, että riippumatta kellonajasta useat kiinteistöt keskustan alueella ovat usein vierailtuja.



Kuva 52: Kolmiulotteinen aika-avaruuskuutiot eri kellonaikoina, kuvat isompana liitteessä 7.






Kuva 53 esittää muutostrendiä kellonajan suhteen, jonka tulos on selkeä laskeva trendi, poikkeuksena yhden tienpätkän alueella lännessä. Tulos viittaa siihen, että työt tehdään useammin aamupäivällä ja päivällä ja viimeisellä työtunnilla seurantalaitte on laitettu pois päältä, eli ei ole aloitettu uusia työtehtäviä.



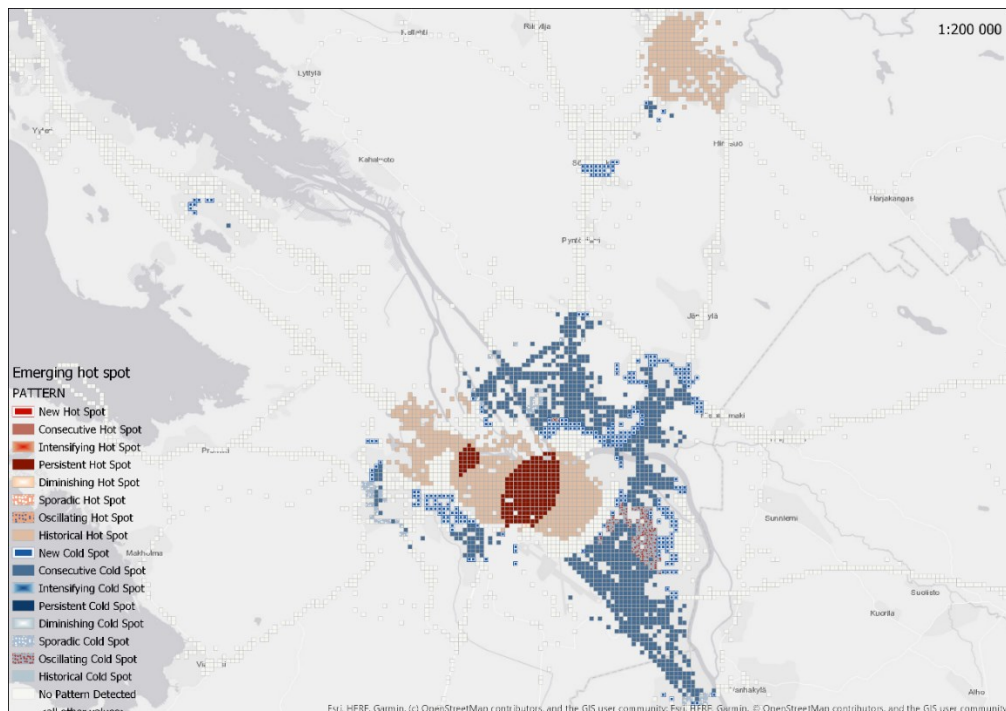
Kuva 53: Muutostrendi kellonajassa klo7-15 seurannassa.

Kuva 54 esittää kuuma- ja kylmäpisteanalyysin tuloksia. Tuloksia voidaan tulkita alla olevan taulokon (Taulukko 5) avulla.

*Taulukko 5: Kellonajan mukaan tehty tulkinta kuuma- ja kylmäpisteanalyysin tuloksista.*

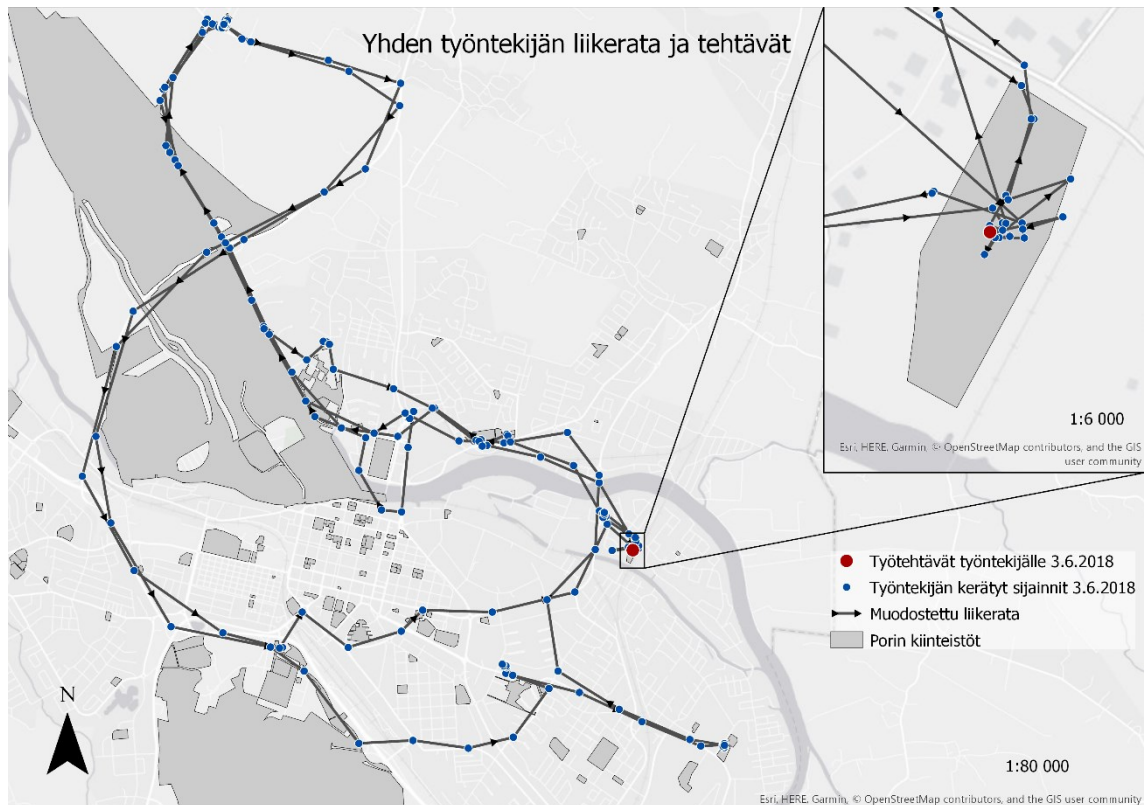
Symboli	Selite	Tulkinta aineistolle
	Persistent hot spot	Suosittu riippumatta kellonajasta
	Historical hot spot	Aamulla ja päivällä suosittu, mutta viimeinen tunti harvinainen
	Oscillating cold spot	päivällä suosittu, viimeinen tunti harvinainen
	Consecutive cold spot	Päivästä iltapäivään erittäin harvinainen, muuten vaihteleva.
	New cold spot	Viimeinen tunti harvinainen, muuten vaihteleva.

Taulukon ja kartan avulla voidaan tulkita, että keskustan alueella on suosituinta tehdä töitä kellonajasta riippumatta. Heti keskustan ympärillä taajama-alueella (ja Noormarkun alueella) on myös hyvin suosittua työskentelyaluetta, mutta viimeinen työtunnin aikana alue on harvinaisemmin vierailtua. Ennen kylmäpistealueita, on puskurialue, joka ei ole tilastollisesti merkittävää aluetta. Melkein kaikki kylmäpistealueet ovat ”Consecutive cold spot” -alueita, eli aamulla ja päivällä harvinaista aluetta ja muuten ei ole merkittävää vaihtelua. Tämä alue, voisi viitata siihen, että suurin osa huollettavista kiinteistöistä on tämän alueen ulkopuolella, enemmän kaupungin keskustassa. Iltapäivällä, kun työvuorot loppuvat, alkaa liikennevirta suuntaamaan ulospäin keskustasta, joillakin on saattanut jäädä seurantalaitte päälle, mutta ei niin merkittävästi, että olisi syntynyt kuumapistealueita.



*Kuva 54: Kuuma- ja kylmäpisteanalyysi kellonajan mukaan klo 7-15.*

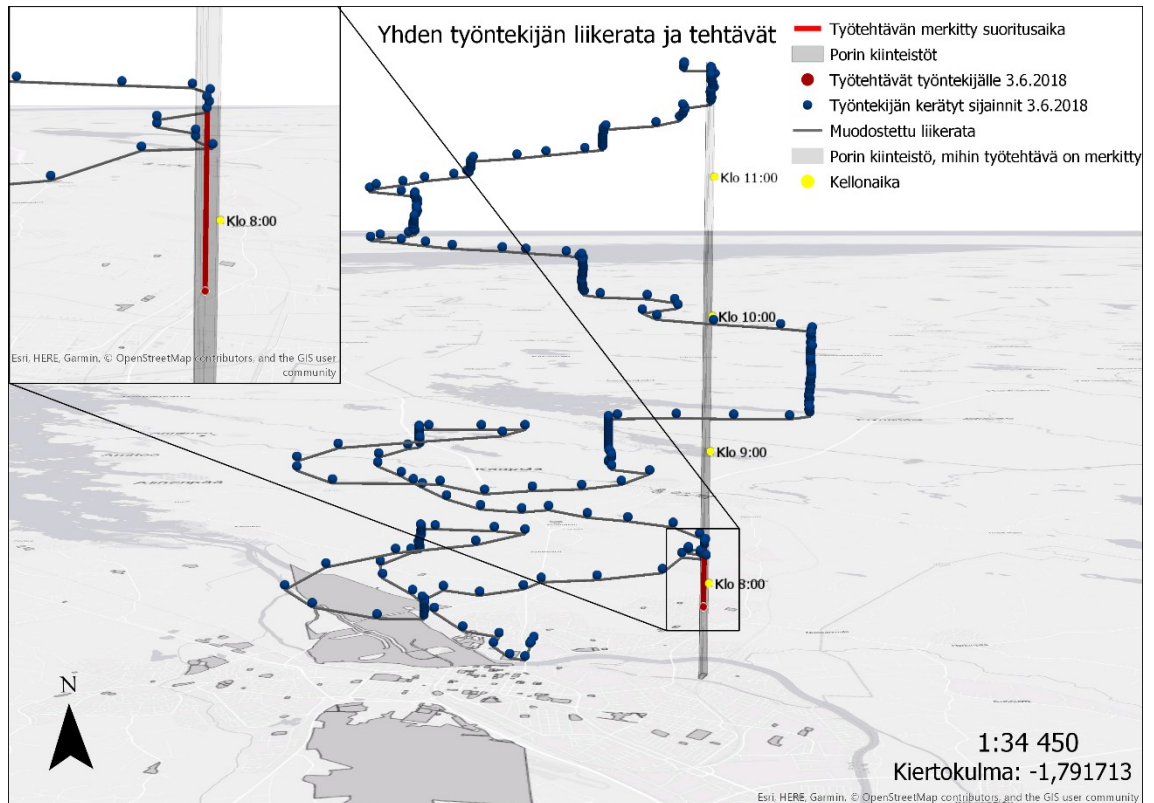




Kuva 55: Yksittäinen liikerata ja työtehtävät

Kartalla (Kuva 55) näkyy yksi satunnaisvalinnalla valittu liikerata, joka koostuu yhden päivän ja yhden työntekijän pisteistä aikajärjestyksessä. Liikerata on visualisoitu mustalla viivalla, missä nuoli kertoo liikeradan suunnan ja aikajärjestyksen. Liikeradasta näkee, että kerättävät pisteet seuraavat verrattain hyvin tieverkkoa, ja että päällekkäiset reitit eivät ole täydellisesti päällekkäin, mutta noudattavat samaa rataa. Lisäksi tehtävänannon läheisyydessä on isompi pistekertymä samalla alueella ja kiinteistölle. Kartasta voidaan tulkita, että työntekijä on käynyt suorittamassa työtehtävän, mutta useat samalla kiinteistöllä olevat saapumiset ja poistumiset viittaavat siihen, että kiinteistöllä on käyty useamman kerran.

Kun liikerataa tarkastellaan kolmiulotteisesti, kartan tulkinta muuttuu ja aikaulottuvuus saa enemmän painoarvoa päällekkäisyyden analysoimisessa. Kartassa (Kuva 56) on sama liikerata, esitetty aika-avaruuskuution omaisesti niin, että kolmas ulottuvuus on aika sekunteina. Jokainen sininen pallo on kerätty piste ja liikerata on generoitu niiden välille. Punainen piste ja viiva kertoo, missä ja milloin tehtävänanto on annettu työntekijälle ja milloin se on suoritettu. Kartasta erottuu selkeämmin paikallaanolo hetket, mitkä ovat kohtisuorassa maata vasten olevat viivasegmentit liikeradasta. Kun liikerataan on yhdistetty vierailtava kiinteistö ja tehtävänanto, päivän kulku selkeytyy. Kartasta voi tulkita selkeästi, että työntekijä on käynyt kahdesti kiinteistöllä. Ensimmäinen käynti liittyy tehtävänantoon, sillä työntekijä on merkinnyt tehtävän alkaneeksi, kun hän alkoi siirtymään kiinteistöä kohti ja suoritetuksi, kun hän poistui kiinteistön alueelta. Lisäksi työntekijä on liikkunut kaksi kertaa saman reitin: ennen ja jälkeen annetun tehtävänannon. Tämä reitti näkyy selkeämmin kaksiulotteisessa kartassa (Kuva 55).



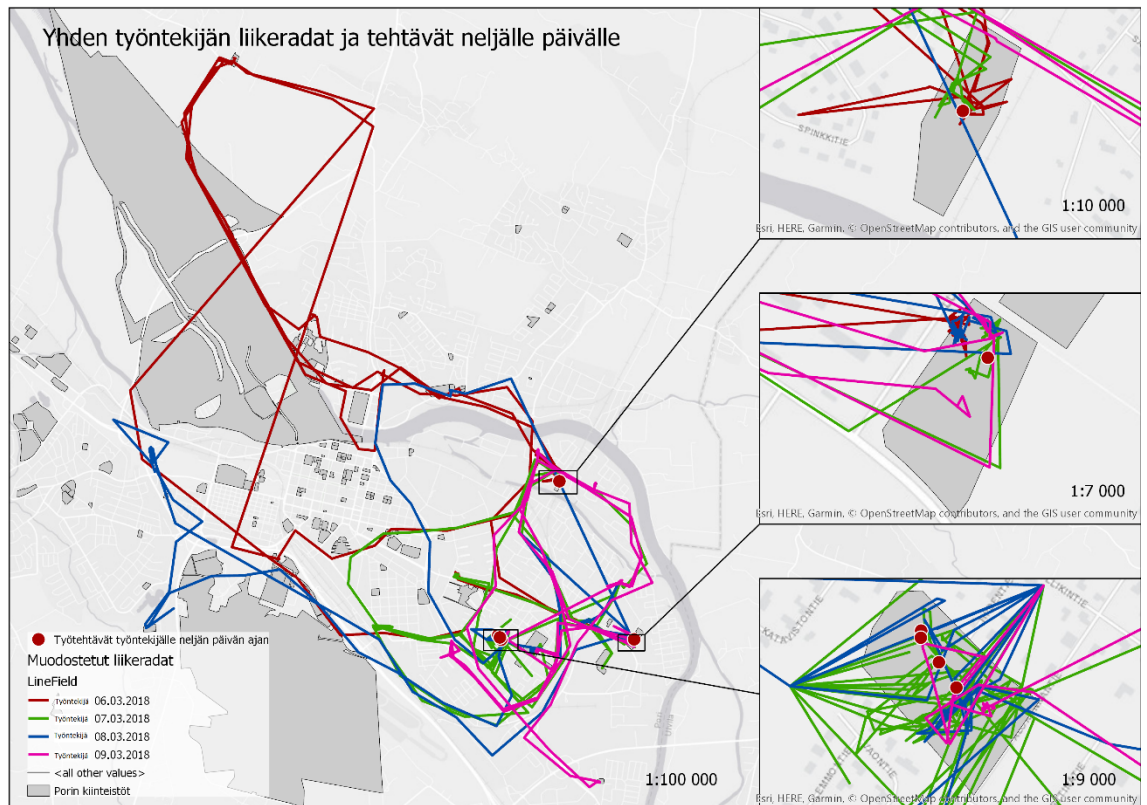
Kuva 56: Yksi liikerata kolmiulotteisena, kolmas ulottuvuus on aika

Liikeratoja on niin monta, että niitä ei voi tulkita yksi kerrallaan, eikä sen tarkoitus ole seurata yksittäisiä työntekijöitä, vaan saada analysoitavaa aineistoa, millä voidaan laskea muutosta tai tilastoja kokonaiskuvasta kiinteistöhuollossa. Toisaalta, kun tarkastellaan vain neljää liikerataa peräkkäisiltä päiviltä ja samalta työntekijältä, aineistoa on jo huomattavasti vaikeampaa tulkita. Aineistosta ei ole siivottu häiriöpisteitä, vaan liikeradat sisältävät vielä kaikki keräyspisteet. Tämän takia, pohjoisosassa karttaa näkyy poikkeama tieverkosta, kun liikerata jatkaa kiinteistöjen yli. Kiinteistöt, missä on tehtävänantoja ovat selkeästi myös suosittuja sijainteja, mutta aineiston tarkkuus ei riitä viivasegmenttien pysymiseen kiinteistön sisälle. Kun liikeradat siivotaan ja visualisoidaan luokittelun mukaan päivämäärän sijaan, huomataan katkoksia liikeradassa sekä luokittelun tarkkuus suhteessa tieverkkoon ja kiinteistöihin. Kuva 58 esittää aineiston esikäsittelyn, segmentoinnin ja luokittelun tuloksia pienemmällä aineistotokseilla.

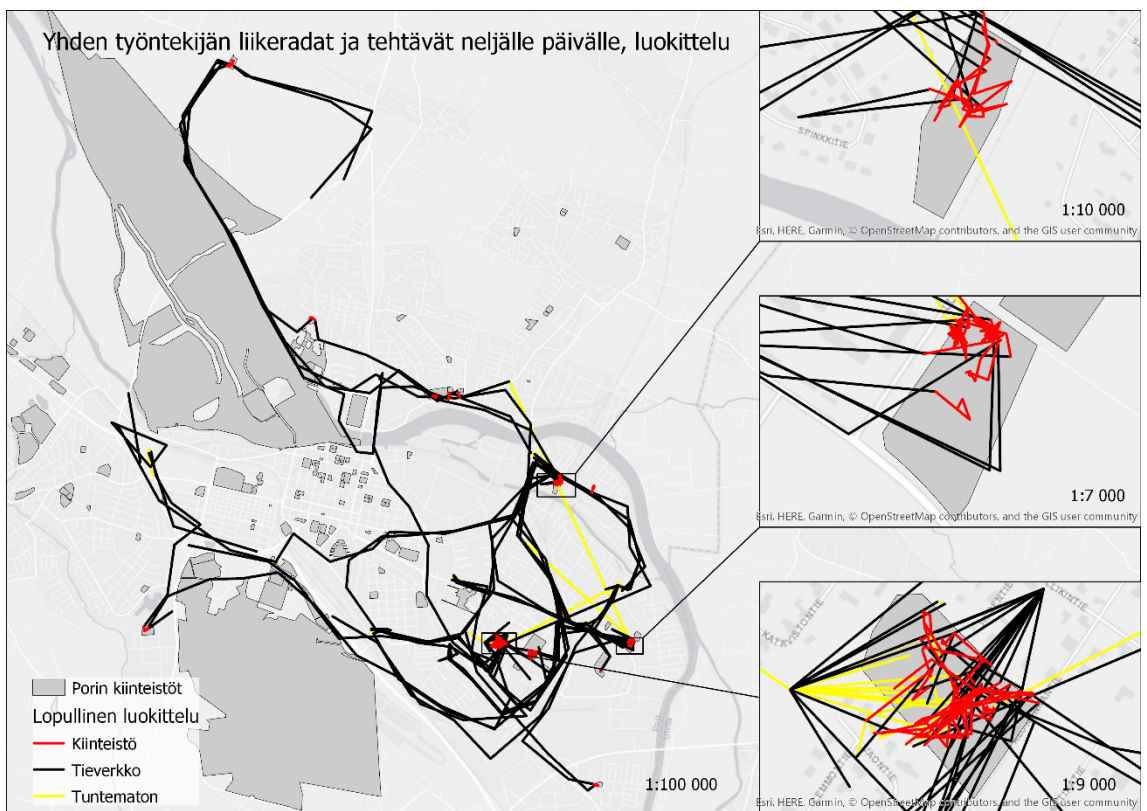
Kartat (Kuva 59 ja Kuva 60) esittävät samoja aineistoja, mutta ovat kolmiulotteisesti esitetty. Mittakaava on pienempi ja kiinteistöillä oleminen on visualisoitu pystyviivalla. Kuva 60 osoittaa, että luokittelu on onnistunut suurimmalle osalle segmenteistä ja tuo lisäymmärrystä aika-avaruusvisualisointiin esittämällä myös oletettua kulkumuotoa.

Isompi aineisto-otos on mahdollista visualisoida samoilla menetelmillä (Kuva 61 ja Kuva 62), mutta tulos ei ole kovin luettava staattisena karttana. Liikeratojen visualisointi tässä laajuudessa tuo esiin isoimpia ilmiöitä, eli kiinteistöjen kohdalle muodostuu pylväsmäisiä visualisointeja ja osa tuntemattomista liikeratasegmenteistä on samanlaisia pylväitä, mikä viittaa siihen, että kyseiset tuntemattomat sijainnit ovat joitain kiinteistöjä, jotka eivät ole tämän tutkimuksen piirissä.

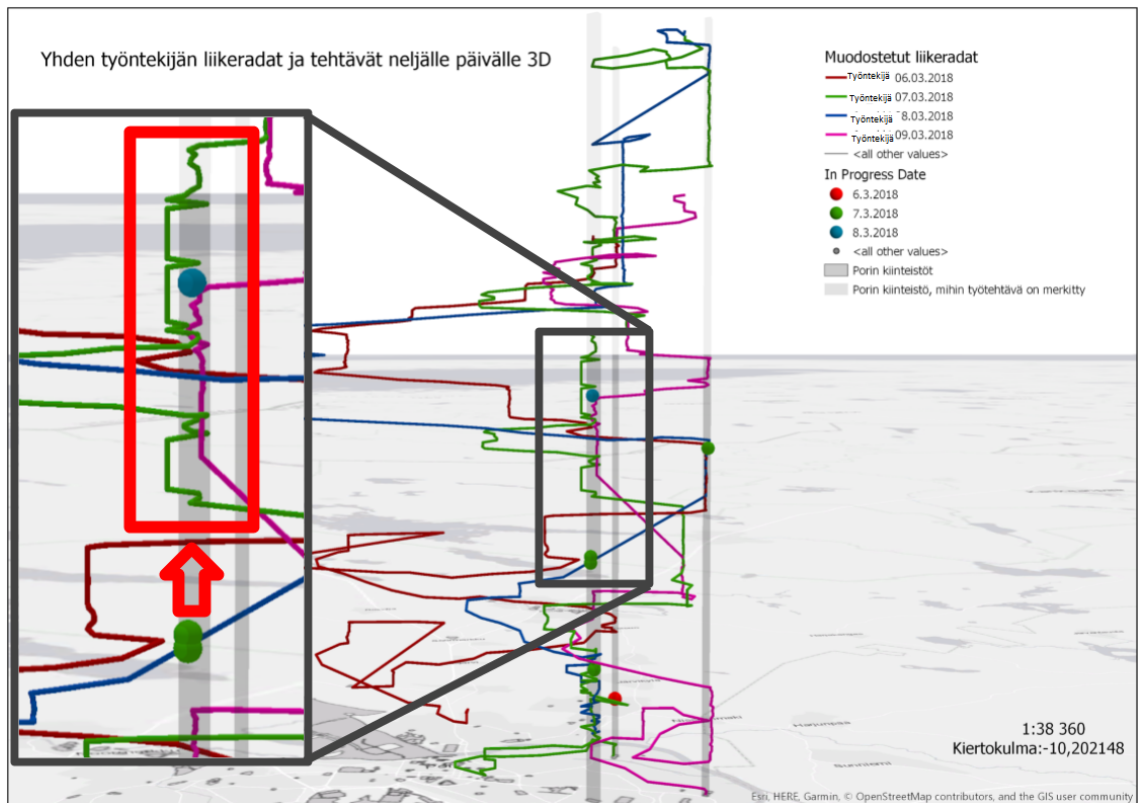




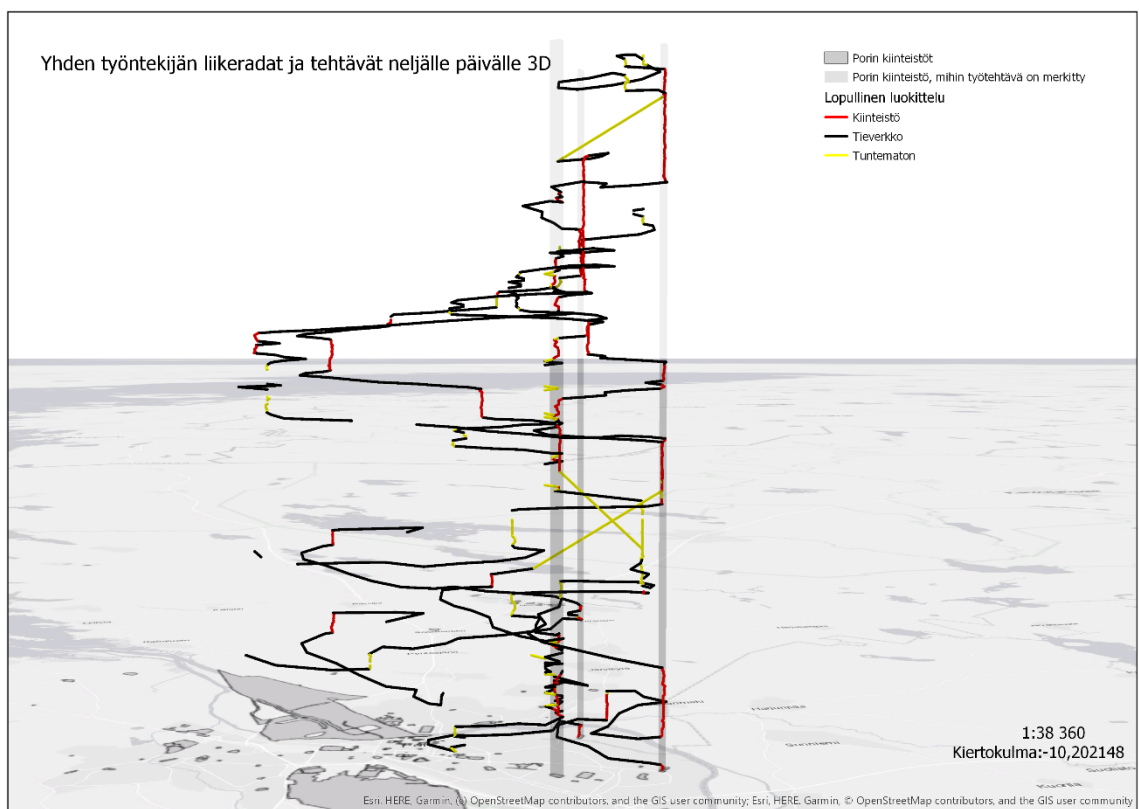
Kuva 57: Neljän päivän aineisto



Kuva 58: Neljän päivän aineisto siivottuna ja luokiteltuna

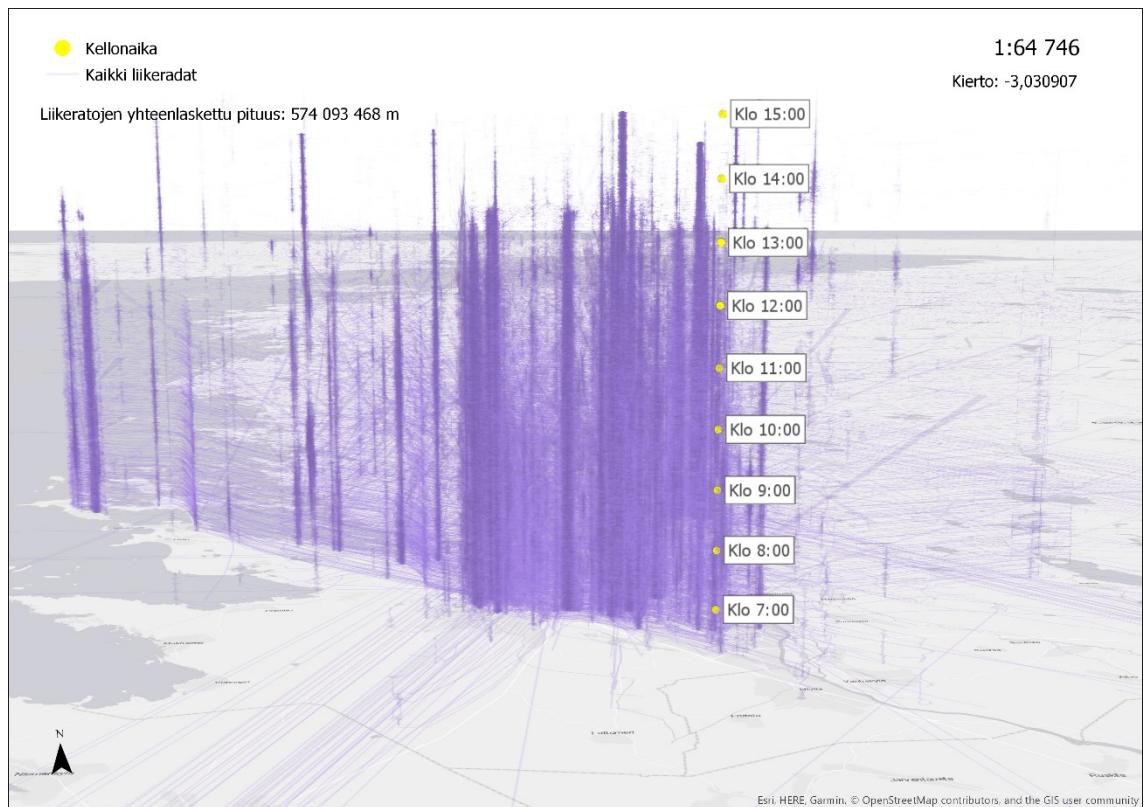


Kuva 59: Neljän päivän aineisto kolmiulotteisesti

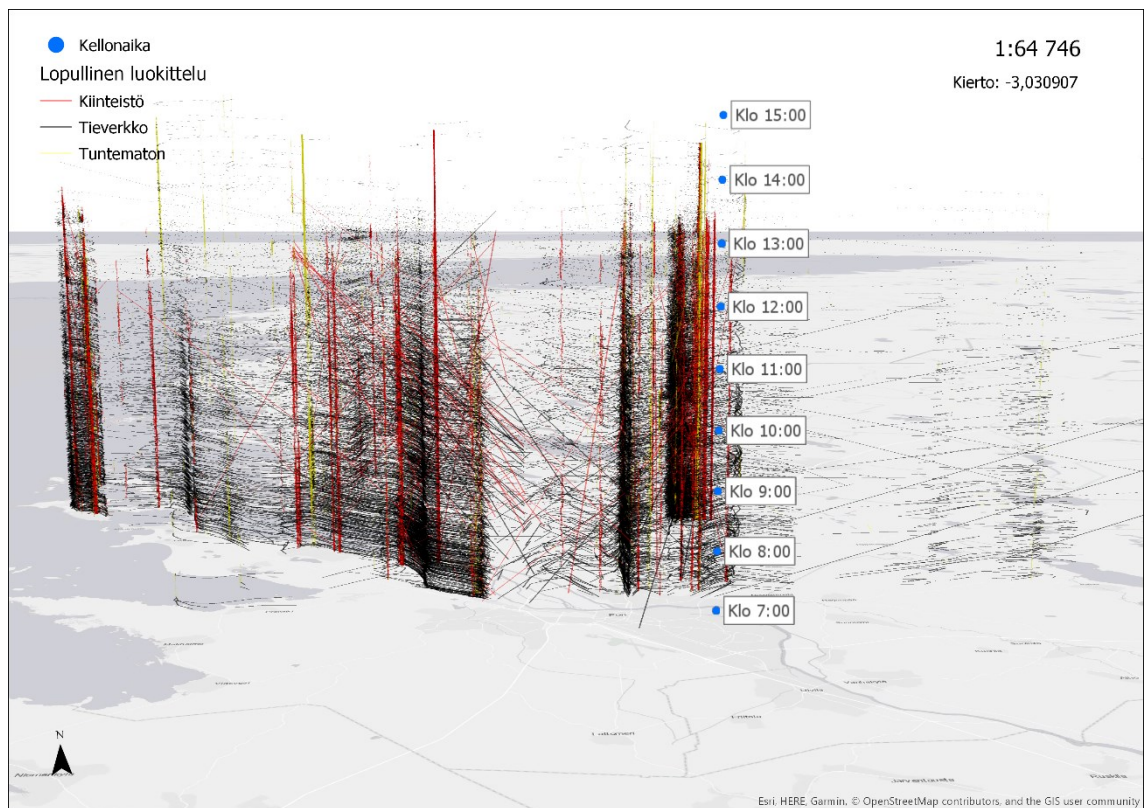


Kuva 60: Yhden työntekijän neljän päivän liikeradat siivottu ja luokiteltu





Kuva 61: Kaikki liikeradat kolmiulotteisesti



Kuva 62: Kaikki liikeradat luokiteltuna kolmiulotteisesti (alle 40 m/s)

## 5 Tulosten arviointi ja vertailu

Tässä luvussa jatketaan tiedon löytämisen prosessia vaiheilla 8 ja 9. Eli arvioidaan saatuja tuloksia ja tulkitaan niitä tiedon löytämisen prosessin tavoitteiden valossa. Arvioinnissa otetaan huomioon aineiston käsittelyn vaikutukset algoritmin tuloksiin. Lisäksi tulkitaan tuloksien hyödyllisyys ja ymmärrettävyys.

Lisäksi pohditaan, miten voidaan hyödyntää löydettyä tietoa ja ymmärrystä. Hyödyntäminen voi tarkoittaa tuloksien visualisointia ja integrointia dynaamiseen ja tuotannossa olevaan aineistoon. Tässä viimeisessä vaiheessa (9) saattaa olla haasteista, sillä löydetty tieto ja mallit sovelletaan käytäntöön, eikä vain staattiseen otokseen aineistosta.

Porin kaupunki arvioi tulokset ja alkuperäiset kommentit löytyvät liitteestä 8.

### 5.1 Tulosten arviointi

Ensimmäisenä arvioidaan saatuja tuloksia ja tulkitaan niitä tiedon löytämisen prosessin tavoitteiden valossa. Ensimmäinen tutkimuskysymys oli saada tekstiviestin lähettäminen automatisoitua, kun tehtävänannon tila muuttuu. Tuloksen soveltaminen käytäntöön on vielä kesken, sillä se vaatii työntekijöiden työtavan muutosta, mikä on esitetty tuloksissa. Tutkimuskysymykseen saatiin kuitenkin tuloksia niin, että toimintamalli ja automatiikka toimii sähköpostin lähettämiseen testiaineistolla.

Prosessi perustuu aikaleimoihin, jotka päivittyvät, kun työntekijä päivittää työtehtävän tilan. Tämän takia prosessissa on paljon toistoa. Prosessissa piti valita lisätäänkö aineistoon uusi kenttä, jota päivitetään vai tehdäänkö erilliset tekstiviestit jokaista tilaa varten. Prosessissa päätettiin käsitellä aineistoa, sillä tekstiviestin konfiguroimista liittymän kanssa ja tekstin sisältöä on helpompi muuttaa. Lopullisessa prosessissa voidaan näin helpommin päättää, mitä tietoja sisällytetään tekstiviestiin. Esimerkiksi sisältyykö työntekijän muistiinpanot vai ei.

Suurin epävarmuus on käyttäjä sillä, jos puhelinnumero kirjoitetaan väärässä muodossa, niin sitä ei voi ilmoittaa tehtävänannon kirjaajalle, sillä sovellus, millä tehtävänanto kirjataan, ei tunnista puhelinnumeroja. Sähköpostin käyttäminen olisi varmempi tapa, sillä sovellus tunnistaa sähköpostit ja pyytää käyttäjää korjaamaan osoitteen, jos se ei sisällä tarvittavia osia (muun muassa @-merkki).

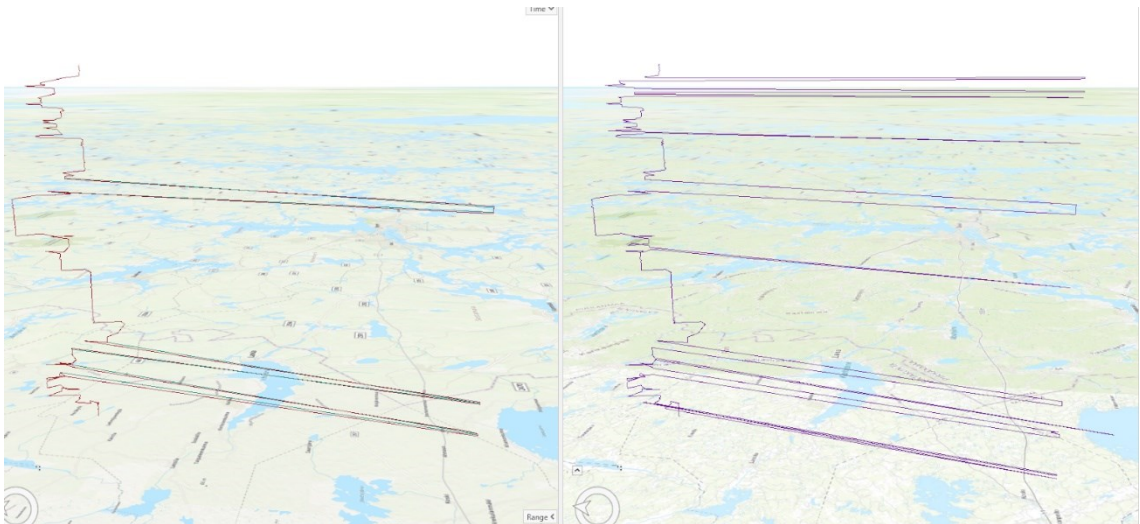
Uusi järjestelmä aiheuttaa myös rajoituksia ja uusia sovelluksia käyttöön tehtävänantojen kirjaajille. Tällä prosessilla pitää käyttää uutta sovellusta (GeoForm), sillä se mahdollistaa lisäkenttien täyttämisen, mitä aiempi sovellus ei mahdollistanut (Workforce for ArcGIS). Aiempi sovellus kuitenkin jää käyttöön, sillä siinä on parempi käyttöliittymä työtehtävien jakamiseen tietyille työntekijälle.

Porin kaupunki ottaa toteutusmallin testikäyttöön ja uskoo tuloksien parantavan asiakastytytyvääisyyttä (liite 8).

Toinen tutkimuskysymys oli hyödyntää kerättyä aineistoa mahdollisimman paljon kiinteistöhuollon parantamiseksi. Porin kaupunki kokee kaiken tiedon hyödylliseksi ja

erityisen hyödylliseksi nostettiin Operation Dashboard -reaaliaikaiset näkymät ja luokittelun mahdollistamat laskelmat. Luokittelun mahdollistavat laskelmat olivat ajomatkan ja työajan vertailu eri ajanjaksoilta. Näiden tuloksien soveltaminen reaaliaikaiseen aineistoon ei ole mahdollista, mutta työnkulku niiden tekemiseen aineistotoksen avulla on olemassa ja käytettävissä tulevaisuuden aineistoissa. Lisäksi Porin kaupunki haluaa, että tehtävänannot ovat osa rutiinitehtävien hoitoa, jonka seurauksena myös rutiinityöt merkitään samaan järjestelmään. Näin saadaan tämän työn analysointimenetelmiä hyödynnettyä laajemmin työnseurantaan ja pitempiaikaiseen tutkimukseen siitä, miten käytännön tehtävien tekemisen muutokset vaikuttavat reitteihin ja työaikoihin sekä löydetään poikkeamia käyttäytymisen malleja tai sijainteja (Liite 8).

Isoimmat epävarmuudet ovat aineiston esikäsittelyssä, missä suodatettiin häiriöpisteitä. Häiriöpisteeksi luokittelussa oli paljon epävarmuutta ja olisi vaatinut enemmän iterointikertoja, jotta tulokset olisivat ilman kohinaa (Kuva 63). Tämä olisi kuitenkin vaatinut ohjelmointia, joka ei kuulunut tämän tapaustutkimuksen aiheisiin, sillä tässä tapaustutkimuksessa käytettiin ainoastaan ArcGIS-tuoteperheeseen kuuluvia tuotteita sellaisenaan.



*Kuva 63: häiriöpisteiden siivoaminen, kolmen iterointikierroksen jälkeen ei ole vieläkaan täysin kohinaton.*

ArcGIS-tuotteiden rajoituksen takia myös klusterointianalyysit jäivät ainoastaan pistetasolle tai visualisointimenetelmin toteutettaviksi, sillä ArcGIS-tuotteet eivät tue viivasegmenttien käyttämistä klusterointialgoritmien ajossa.

Klusterointimenetelmien tulokset ja niiden hyödyllisyys korostuu, kun halutaan vertailla aineistojoukkoja tietyllä aikavälillä ja tietyillä työntekijöillä. Klusterointialgoritmit korostavat sijainteja, missä on vierailtu eniten. Klusterointimenetelmät ovat suuresti riippuvaisia annetuista parametreista ja tulokset vaihtelevat suuresti niistä johtuen, joten vertailussa on hyvä pitää samat parametrit.

Näillä parametreilla sopivin algoritmi on HDBSCAN, sillä se löysi eniten klustereita annetuilla parametreilla ja ne korreloivat vahvasti olemassa oleviin kiinteistöihin. HDBSCAN optimoi parametrit pitkälti itse, joten käyttäjän vaikutus on pieni, verrattuna DBSCAN ja OPTICS algoritmeihin. Toisaalta, DBSCAN ja OPTICS tuloksien klusterit ovat pinta-alaltaan pienempiä kuin HDBSCAN tuloksen klusterit.

Vertailussa olevien ajanjaksojen, marraskuu 2017 ja marraskuu 2018, käytettävien aineistojen kokoero on huomattava: 2017 aineistossa on 68000 pistettä ja 2018 on 41000 pistettä. Tämän takia 2017 analyysistä löytyi enemmän klustereita HDBSCAN ja OPTICS analyysissä, mutta DBSCAN analyysissä niitä löytyi vähemmän.

Porin kaupunki hyödyntää tuloksia muun muassa uusien tukikohteiden suunnitteluun, taukopaikkojen paljastamiseen ja lisäresursoinnin suunnitteluun. Lisäksi paljon työtä vaatineiden kohteiden tarkempi analyysi voisi johtaa järjestelmien tai koko talon peruskorjaukseen (Liite 8).

Luokittelun tuloksien hyödyllisyys on suuri, sillä luokittelun ansiosta ajometrien määrä voitiin laskea ja vertailla kuukausien välistä muutosta. Muutos oli suuri marraskuu 2017:ta ja marraskuu 2018:ta välillä, sillä ajokilometrit pienenivät 37% 14 700 kilometristä 9 200 kilometriin. Näissä tuloksissa ei ole otettu huomioon toimintatapoja. Toimintatavat voivat vaikuttaa suuresti lopputulokseen, sillä seurantalaitte pitäisi olla päällä vain näitä työtehtäviä hoidattaessa. Klusterointianalyysit osoittavat, että työntekijät pitivät seurantaan päällä myös työtehtävien ulkopuolella, mikä vääristää tuloksia. Vuoden aikana tulos muuttuu muuan muassa siksi, että seurantalaitetta on opittu käyttämään paremmin. Luokittelun avulla voidaan myös määritellä työtuntien määrää niin ajon aikana kuin kiinteistöllä halutulla ajanjaksolla. Tulokset osoittavat, että noin puolet työajasta on huollettavilla kiinteistöllä ja 26 % matkustamisessa. Loput työajasta on seurattavien työtehtävien ulkopuolista toimintaa sekä luokittelematonta aineistoa.

Porin kaupunki hyödyntää tuloksia kiinteistöhoidon organisoinnin muutoksen vaikutuksien löytämiseen ja tiedottamiseen, sillä luokittelun ansiosta näkyy selkeä muutos ajokilometreihin ja kiinteistöllä työajan viettämiseen. Tulokset vaativat lisätutkimusta työntekijöiden kesken, mistä tulokset johtuvat ja onko vaikutus halutunlainen. Lisäksi visualisointimenetelmät olivat selkeitä ja tarkoituksen mukaisia ja jopa tieteellistä taidetta. (Liite 8.)

Liikeratojen visuaalinen esittäminen on hyödyllistä, sillä symboliikan säätämällä pystyi jo tekemään alustavia tulkintoja suosituimmista paikoista ja eniten käytetyistä reiteistä (Kuva 64). Visualisointia pystyy suodattamaan esimerkiksi tietylle ajanjaksolle tai seurattaville kohteille, minkä jälkeen tehdyt vertailut paljastavat muutokset ajassa tai seurattavien kohteiden välillä. Porin kaupunki hyödyntää yksittäisten työntekijöiden liikeratoja vertailuun ja työtaakan tasapainottamiseen (Liikeradat liitteessä 2). Lisäksi Porin kaupunki haluaa hyödyntää liikeratojen visualisointia laajemmin mm. vakuutustapauksissa tai web-sovelluksina. (Liite 8.)





Kuva 64: Kaikki liikeradat visualisointina niin, että klusteroituminen on havaittavissa.

Operation Dashboard -sovelluksen hyödyllisyys perustuu kokonaiskuvan tarkasteluun ja sovelluksen interaktiivisuuteen. Lisäksi sovelluksen informaatiota ja sisältöä on helppo muokata tarkoituksen mukaan, joten tuloksissa esitetty sovellus ei ole viimeinen kehityksen muoto, vaan tämän hetkinen versio. Sovelluksen tarkoitus on myös herättää mielenkiintoa tarkastelemaan tietoja yhteenvetonäkymillä ja synnyttämään ideoita uusille näkymille.

Sovelluksen heikoin kohta on, että siihen on laitettu liikaa tietoa yhteen näkymään, sillä saman tiedon olisi voinut esittää useammalla sovelluksella. Sovelluksen laajuuden valitsemisessa pitää valita sovelluksien määrän ja selkeyden välillä, sillä loppukäyttäjä ei halua vaihtaa sovellusta kesken työnkulkua, mutta ei halua ylimääräistä tietoa sovellukseen.

Porin kaupunki käyttää jo aktiivisesti Operation Dashboard -sovelluspohjaa, mutta nämä tulokset toivat uusia tapoja hyödyntää sovellusta, sillä he eivät ole laittaneet kahta karttaa samaan sovellukseen tai suodattaneet tehtävänantoja niiden tilan perusteella. Porin kaupunki ottaa sovelluksen käyttöön. (Liite 8.)

Aika-avaruuskuutiossa olevien liikeratojen visualisointi toi uusia ulottuvuuksia aineiston tulkintaan, mutta hyödyllisyyttä löytyi vain siivotusta ja erittäin pienestä otoksesta tarkasteltuna. Loppukäyttäjälle miellyttävien visualisointi oli yksittäisen liikeradan esittäminen aika-avaruuskuutiossa, sillä liikeradat risteävät voimakkaasti, eivätkä noudata muuta kuin tunnettua tieverkkoa, joten yleiskuvan saaminen koko aineistosta on hyvin vaikeaa staattisesta kuvasta. Dynaamisena esitystapana aika-avaruuskuutio on raskas tietokoneelle esittää, mutta siitä pystyy ymmärtämään visuaalisesti paljon, kun käyttäjä saa itse valita katselukulman.

Liikeratojen z-arvon määrittäminen olisi voitu tehdä toisella tavalla, sillä nykyinen sekuntien määrä, nostaa liikeradat jopa 28 kilometrin korkeuteen. Tämä aiheuttaa haasteita visualisoinnin ymmärtämiseen ja käsittelyyn ArcGIS Pro -ohjelmassa. Sekunnit valittiin yksiköksi z-arvoihin, sillä ne olivat kokonaislukuja ja helposti muutettavaksi takaisin minuuteiksi ja tunneiksi. Toisaalta sekunneista on vaikea käsittää ajankulkua pelkkänä numerona, joten jako tunteihin tai minuutteihin desimaalilukuna olisi ehkä toiminut paremmin. Lisäksi liikeradat jäivät segmentoimisen ja siivoamisen jälkeen katkonaisiksi. Katkonaisuus olisi voitu korjata muodostamalla uudet viivasegmentit puuttuvien pätkien väliin ja laskemalla niille uudestaan nopeus ja luokittelemalla aineisto. Tästä olisi kuitenkin syntynyt lisää häiriökohteita, jotka pitäisi siivota ja koko prosessi pitäisi tehdä uudelleen ja uudelleen. ArcGIS-ohjelmisto ei tarjoa tähän automatisoitavissa olevaa ratkaisua ilman ohjelmointia, joten se päätettiin jättää pois aineiston esikäsittelyvaiheessa. Vaikutus lopputulokseen on visuaalinen, mutta myös osittain tilastollinen, sillä puuttuvat segmentit vähentävät laskettua työaikaa ja -matkaa, joiden laskelmat perustuvat liikeratojen segmenttien geometriaan.

Porin kaupunki hyödyntää liikeratojen kolmiulotteisen visualisointia työkaluturin seurantaan ja ymmärtämiseen. Liikeratoja tarkastelemalla huomattiin, missä vaiheessa työtehtävä merkitään aloitetuksi ja lopetetuksi ja miten työntekijän sijainti riippuu näistä teoista. Jotta aineiston laatua voidaan parantaa, aineiston keruu pitää yhtenäistää. Tämä tarkoittaa, että työntekijät pitää kouluttaa tarkemmin työtehtävien merkitsemiseen niin, että kaikki tekevät sen samalla tavalla. Tavoitteena on lisätä kiinteistöllä vietettyä aikaa. (liite 8.)

Aika-avaruuskuutiot viikoittain tai kellonajan mukaan esitettynä toivat ison määrän aineistoa paremmin tulkittavaksi verrattuna liikeratojen aika-avaruuskuutiossa visualisointiin. Aika-avaruuskuutioiden rajoituksena oli, että työkalu toimi vain pisteaineistolla, mutta näin pitkällä ajanjaksoilla kuten kuukausi, viikko, päivä tai tunti itse liikeradan reitillä ei ole merkitystä verrattuna, missä ja milloin oltiin eniten ja miten trendit muuttuvat ajanjaksojen yli. Näin ollen tämä ohjelmiston rajoitus ei aiheuttanut suurta menetystä itse analyysissä.

Dynaamisesti käytettynä kolmiulotteiset aika-avaruuskuutiot olivat hyvä visualisointimenetelmä aineiston yleisnäkymää varten ja yksittäisten ajanjaksojen tulkitsemiseen ja vertailuun. Staattisena karttana kolmiulotteinen visualisointi ei toiminut. Sen sijaan kaksikulotteiset trendianalyysivisualisoinnit toivat eniten hyötyä aineiston tulkitsemiseen, sillä trendianalyysit vertailevat eri ajanjaksoja ja löytävät sijainteja, joissa on merkittävä muutos ajan suhteen, esimerkiksi alueella on kasvava käyntitrendi tai tehtävänantoja. Epävarmuutta lisää työntekijöiden sijaintipisteiden määrän lasku ja sen vaikutus trendianalyysiin. Trendianalyysien tulokset viittaavat, että yleisin trendi on laskeva muutamaa sijaintia lukuun ottamatta, vaikka tehtävänantojen määrä on pysynyt samana. Työtehtäviä tarkasteltaessa analyysit eivät löydä suuria poikkeamia ajan ja paikan suhteen niin kuukausi- kuin viikkotasolla tarkasteltuna.

Porin kaupunki kokee aika-avaruuskuutioanalyysit vaikeaselkoisiksi, erityisesti kolmiulotteisesti, mutta kokee, että analyysit tuovat esiin kokonaiskuva ja muutoksen suuntaa esiin ilman, että seurataan yksittäisiä työntekijöitä. Näin ollen työntekijätkin ymmärtävät paremmin heidän sijaintiseurantansa syitä ja hyödyntämisen mahdollisuuksia. (Liite 8.)



## 6 Johtopäätökset

Pelkän sijainnin ja aikaleiman avulla pystyy analysoimaan erittäin paljon, mutta oikeasti hyödyllisten analyysien löytäminen on vaikeaa. Apua tähän toi KDD, tiedon löytämisen prosessi, sillä sen työjärjestys pakottaa aineiston käsittelijää miettimään tavoitteita ja aineiston hyödynnettävyyttä pitkin prosessia. Tuloksien arviointi ja hyödyntäminen nousee prosessin tärkeimmäksi vaiheeksi. Prosessista poikettiin osittain menetelmän valinnan vaiheessa, sillä silloin ei rajoitettu vain yhteen menetelmään. Toinen tulkinta on, että KDD tehtiin moneen kertaan ja jokaista menetelmää varten erikseen, sillä menetelmiä on vain kolme: klusterointi, luokittelu ja visuaaliset menetelmät.

Hyödyllisimmiksi menetelmiksi ja algoritmeiksi valikoitui klusteroinnissa HDBSCAN, luokittelussa päätöspuu ja visuaaliseksi menetelmäksi aika-avaruuskuutiot viikoittain ryhmiteltynä.

Tutkimus tuotti paljon uusia tutkimuksen ja kehityksen aiheita, sillä lähtöaineisto ja käytössä olevat ohjelmistot tarjoavat paljon erilaisia ratkaisuja. Tuloksista voisi hyödyntää pitemmälle esimerkiksi luokittelua ja ajomatkoja. Ajomatkan ja nopeuden perusteella voitaisiin laskea kiinteistönhuollon aiheuttama polttoainekulutus ja päästöt ja sen avulla aktiivisesti seurata ja pyrkiä vähentämään päästöjä.

Yksi jatkotutkimuksen aihe olisi automatisoida tiedonsiirtoa ja aineiston esikäsittelyä kiinteistönhuollon tapahtumista pitemmälle. Työtehtävien kirjaaminen valmiiksi ja työajan laskenta voitaisiin laskea reaaliaikaisella aineistolla käyttämällä GeoEvent Server -ohjelmistoa, jos aineiston keräystarkkuus on riittävän hyvä. Lisäksi voitaisiin automatisoida liikeratojen muodostaminen ja siivoaminen esimerkiksi päivittäin ajettavalla skriptillä. Esikäsittelyyn voisi lisätä myös sijainnin seurannan ja työtehtävän välisen relaation, jotta liikeradat voitaisiin segmentoida työtehtävän mukaan, mikä voisi olla osana luokittelua. Näin voitaisiin luokitella aineistoa tehtävätyypin perusteella ja tarkastella ajomatkaa ja työaikaa eri tehtävätyyppien välillä. Luokittelun avulla voitaisiin esimerkiksi tutkia kiinteistöjen huoltohistorian eroja ja ennustaa tulevia remontointitarpeita.

Toinen jatkotutkimuksen aihe olisi työnjaon kuormittavuuden seuraaminen ja mittaaminen. Työntekijän seuraamista voitaisiin hyödyntää työntekijän hyvinvoinnin edistämiseksi ja ennaltaehkäistä tehtäväjonojen syntymistä ja esimerkiksi optimoida työtehtävien jakamista reaaliaikaisen sijainnin perusteella.

Suurin haaste työssä oli käytettävän ohjelmiston rajoitukset. Kirjallisuuskatsauksessa löytyneet liikeradoille tarkoitetut algoritmit eivät olleet lainkaan saatavilla tai soveltuivat ainoastaan pisteaineiston analysoimiseen. Suurimmat puutteet olivat klusteroinnissa ja aika-avaruuskuutioiden analysoinnissa. Lisäksi reaaliaikaisen tiedon prosessointi palvelimella toi suuria haasteita, kun jatkotutkimuksen aiheeksi jäänyt työajanseuranta reaaliaikaisesti tuotti vaihtelevia tuloksia eikä ollut vielä sovellettavissa käytäntöön.

## 7 Yhteenveto

Tutkimuksen tavoitteena oli hyödyntää kiinteistöhuollon työntekijöiden sijaintia ja työtehtäviä tiedonvälityksen parantamiseksi ja hyödyntää kerättyä aineistoa mahdollisimman paljon erilaisilla menetelmillä ja ArcGIS-ohjelmistoilla.

Tutkimuksen kirjallisessa osiossa tutustuttiin tiedon löytämisen prosessiin ja kolmeen tiedon louhintamenetelmään: klusterointiin, luokitteluun ja visuaalisiin menetelmiin. Tutkimuksen empiirisessä osiossa perehdyttiin tarkemmin lähtöaineistoon ja käytössä oleviin ohjelmistoihin. Empiirisessä tutkimuksessa oli kaksi tutkimuskysymystä:

1. Miten voi automatisoida ilmoituksen lähettäminen työtehtävän tilan muuttuessa?
2. Miten Porin kaupungin kiinteistöhuollossa kerättävää aineistoa voidaan hyödyntää kiinteistöhuollon toiminnassa?

Empiirisen tutkimuksen tuloksena oli ensimmäistä tutkimuskysymystä varten selvitys, suunnitelma muutoksesta ja testiaineistolla todennettu malli tekstiviestien lähettämiseen. Toisen tutkimuskysymyksen tuloksena oli esikäsitelty ja luokiteltu liikerata-aineisto, pisteaineistosta klusterointituloksia, aika-avaruuskuutioanalyysia niin viikko kuin tuntikohtaisesti ja analysoitua tietoa kiinteistöhuollon työtehtävien suoritustavoista ja paikoista.

Tutkimuksen tuloksia voi hyödyntää muissa projekteissa, missä kerätään sijaintia ja ajansuhteen. Yhteensopivuus on parhaimmillaan, jos projektissa käytetään Workforce for ArcGIS -sovellusta työtehtävien jakamiseen.

## Lähdeluettelo

- Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. 1999. *OPTICS: ordering points to identify the clustering structure*. ACM Sigmod record. Vol. 28, No. 2. s. 49-60.
- Bédard, Y. & Han, J. 2009. *Fundamentals fo Spatial Data Warehousing for Geographic Knowledge Discovery*. Taylor & Francis.
- Bennet, L et al. 2017. *Spatial Data Mining II: A Deep Dive into Space-Time Analysis*. Esri. Saatavissa:  
<https://www.esri.com/videos/watch?videoid=0aV6HHwJuo4&title=spatial-data-mining-ii-a-deep-dive-into-space-time-analysis> [Viitattu: 3.3.2019]
- Birant, D. & Kut, A. 2007. *ST-DBSCAN: An algorithm for clustering spatial-temporal data*. Data and Knowledge Engineering 60, 1, s. 208–221.
- Campello, R. J., Moulavi, D., & Sander, J. 2013. *Density-based clustering based on hierarchical density estimates*. Pacific-Asia Conference on Knowledge Discovery and Data Mining. s. 160-172.
- Devillers, R., Gervais, M., Jeansoulin, R. & Bédard, Y. 2002. *Spatial data quality: From metadata to quality indicators and contextual end-user manual*. OEEPE/International Society for Photogrammetry and Remote Sensing (ISPRS)
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. 1996. *A density-based algorithm for discovering clusters in large spatial databases with noise*. Kdd. Vol. 96, No. 34. s. 226-231.
- Esri Deutschland. 2019. *Karten entdecken, verwenden, erstellen und teilen – mit jedem Gerät, an jedem Ort, zu jeder Zeit*. Saatavissa:  
<https://www.esri.de/produkte/arcgis/plattform> [Viitattu: 8.3.2019]
- Esri Inc. 2018. *Automate and script Workforce*. Saatavissa:  
<https://doc.arcgis.com/en/workforce/android-phone/help/workforce-schema.htm> [Viitattu: 1.3.2019]
- Esri Inc. 2019. *ArcGIS Online Create maps, scenes, and apps*. Saatavissa:  
<https://doc.arcgis.com/en/arcgis-online/create-maps/create-maps-and-apps.htm> [Viitattu: 28.3.2019]
- Esri Inc. 2019. *ArcGIS Pro tool reference*. Saatavissa: <https://pro.arcgis.com/en/pro-app/tool-reference/main/arcgis-pro-tool-reference.htm> [Viitattu: 28.3.2019]
- Esri Inc. 2019. *ArcGIS Server, What is ArcGIS GeoEvent Server?* Saatavissa:  
<http://enterprise.arcgis.com/en/server/latest/get-started/windows/what-is-arcgis-geoevent-server.htm> [Viitattu: 28.3.2019]

- Esri Inc. 2019. *Create Space Time Cube by Aggregating Points*. Saatavissa: <https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/create-space-time-cube.htm> [Viitattu 20.3.2019]
- Esri Inc. 2019. *Operation Dashboard for ArcGIS*. Saatavissa: <https://doc.arcgis.com/en/operations-dashboard/help/what-is-a-dashboard.htm> [Viitattu: 28.3.2019]
- Esri Inc. 2019. *Workforce for ArcGIS, Create your first project*. Saatavissa: <https://doc.arcgis.com/en/workforce/android-phone/help/create-your-first-project.htm> [Viitattu: 28.3.2019]
- Finlex 13.8.2004/759. *Laki yksityisyyden suojasta työelämässä*. Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/2004/20040759#L5> [Viitattu: 26.1.2019]
- Guptill, S. C. and Morrison, J. L. 1995. *Elements of Spatial Data Quality*. Elsevier Science, New York.
- Hamed, K. 2009. *Exact distribution of the Mann-Kendall trend test statistic for persistent data*. Journal of Hydrology. s.86–94.
- Han, J., Kamber, M. & Pei, J. 2012. *Data mining: Concepts and techniques*. 3rd ed. Burlington, Mass.: Elsevier.
- Han, J., Lee, J. & Kamber, M. 2009. *An Overview of Clustering Methods in Geographical Data Analysis*. Geographic Data Mining and Knowledge Discovery.
- Hägerstrand, T. 1970. *What about people in regional science?* Papers, Regional Science Association, 24, s. 7-21.
- Maimon, O. & Rokach, L. 2010. *Introduction to knowledge discovery and data mining*. Data Mining and Knowledge Discovery Handbook. Eds., 2nd ed. Springer, 2010, s. 1–15.
- Miller, H. & Han, J. 2009. *Geographic Data Mining and Knowledge Discovery*. Chapman and Hall/CRC, Baton Rouge.
- Moen, P. 2013. *Tiedon louhinnan menetelmät. Kurssin keskeisten käsitteiden englanti-suomi-sanasto*. Saatavissa: <https://www.cs.helsinki.fi/u/ronkaine/tilome/sanasto.html> [Viitattu: 13.2.2019]
- Roddick, J. & Lees, B. 2009. *Spatio-Temporal Data Mining Paradigms and Methodologies*. Chapman and Hall/CRC, Baton Rouge
- Tobler, W. 1970. *Computer movie simulating urban growth in Detroit region*. Economic Geography, 46: 2, s. 234–240.
- T. Hägerstrand. *What about people in regional science?* Papers, Regional Science Association, 24, 7-21. 1970.

Veregin, H. 1998. *Unit 100 – Data Quality Measurements and Assessment*. NCGIA Core Curriculum in Geographic Information Science. NCGIA Core Curriculum in GIScience.

Yang, X. et al. 2018. *A Review of GPS Trajectories Classification Based on Transportation Mode*. Vol. 18, Iss. 11. MDPI. Sensors. Saatavissa: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6263992/> [Viitattu: 5.4.2019]

Zaki, M. & Meira, W. 2014. *Data mining and analysis: fundamental concepts and algorithms*. New York: Cambridge University Press. s. 375-340.

Zheng, Y. 2015. *Trajectory data mining: An overview*. ACM Transactions on Intelligent Systems and Technology 6, 3, s. 1–41.



## **Liiteluettelo**

Liite 1. Aineiston automaattisen prosessoinnin kaavio (1 sivu)

Liite 2: Liikeratojen visualisointi (10 sivua)

Liite 3. Klusterianalyysin tulokset yksittäisistä työntekijöistä (3 sivua)

Liite 4. Luokittelun tulokset (3 sivua)

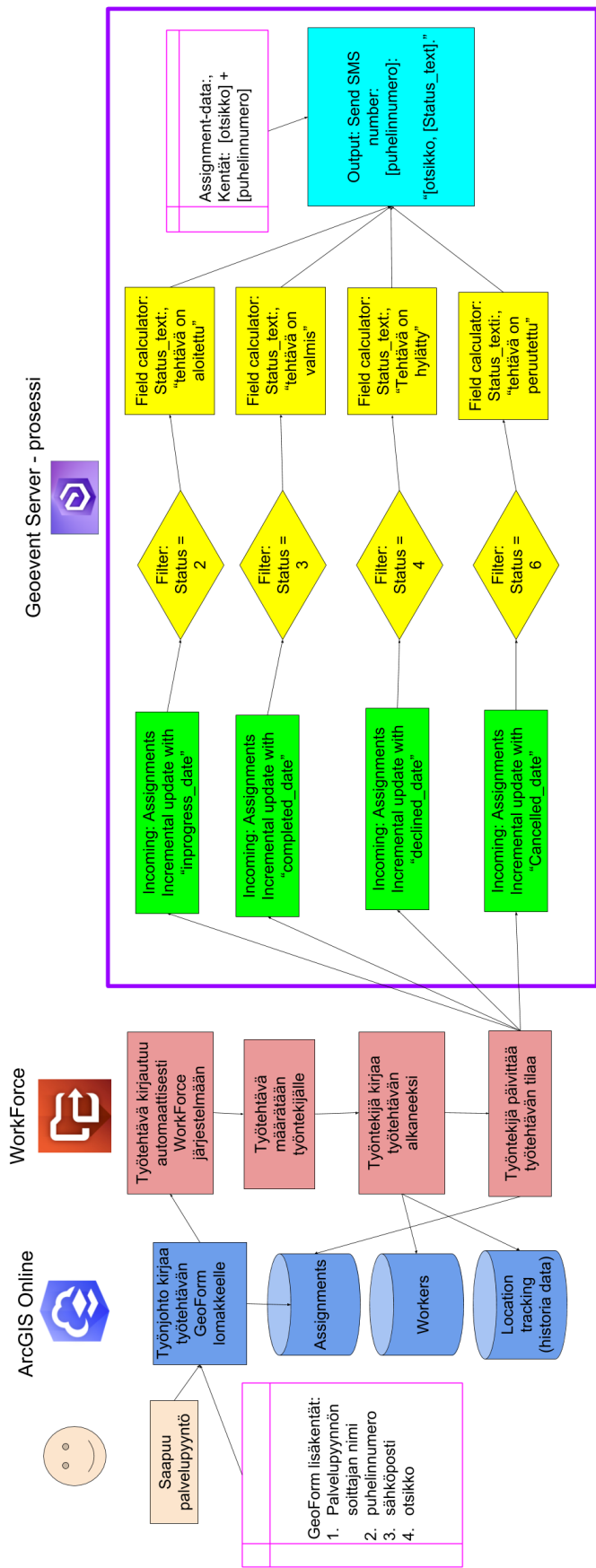
Liite 5. Aika-avaruuskuutioiden trendien analyysitulokset (2 sivua)

Liite 6. Aika-avaruuskuutioiden kuuma- ja kylmäpisteanalyysitulokset (6 sivua)

Liite 7. Aika-avaruuskuutioiden tulokset kellonajan mukaan (7 sivua)

Liite 8. Porin kaupungin kommentit työn tuloksista (2 sivua)

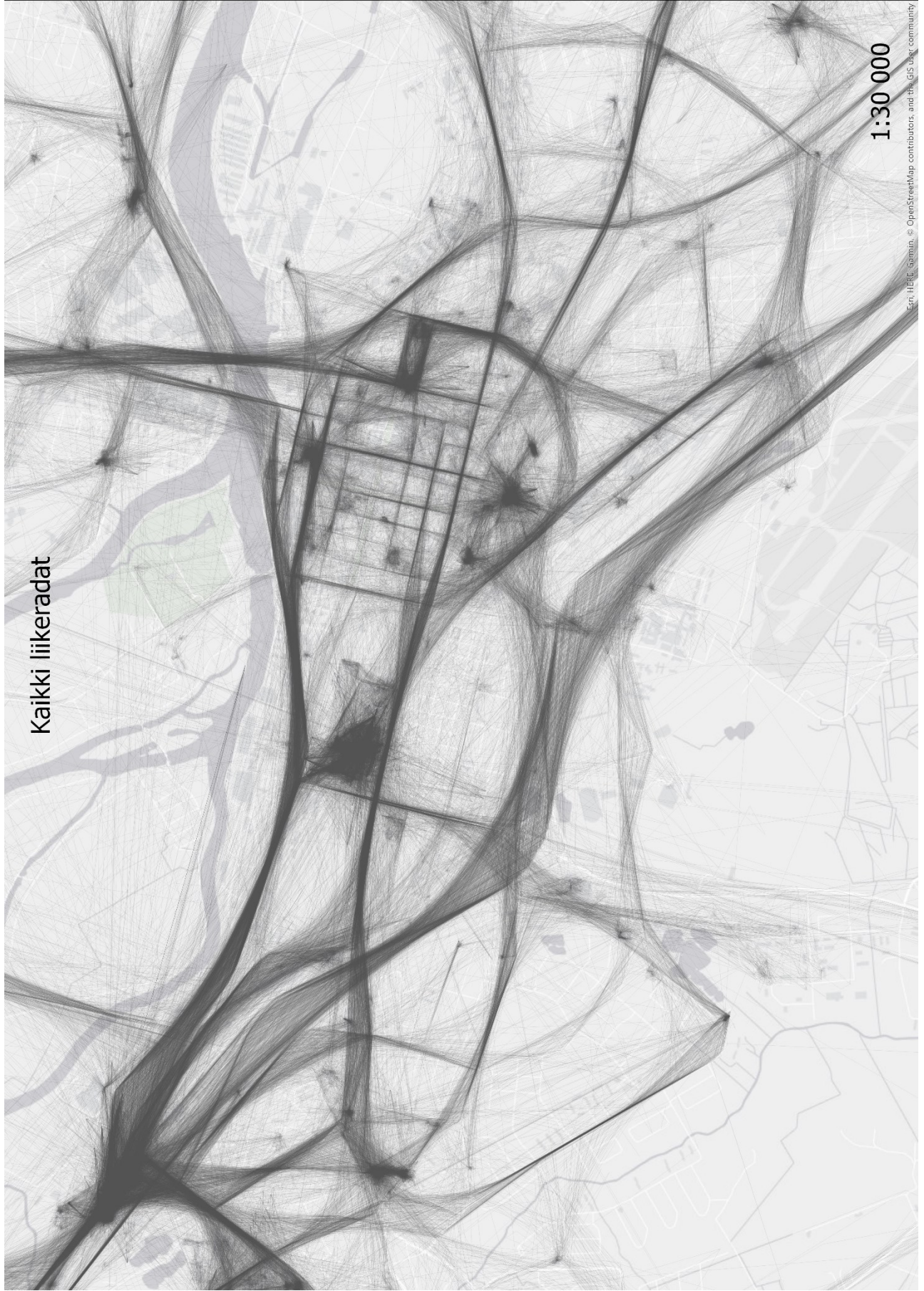
# Liite 1. Aineiston automaattisen prosessoinnin kaavio



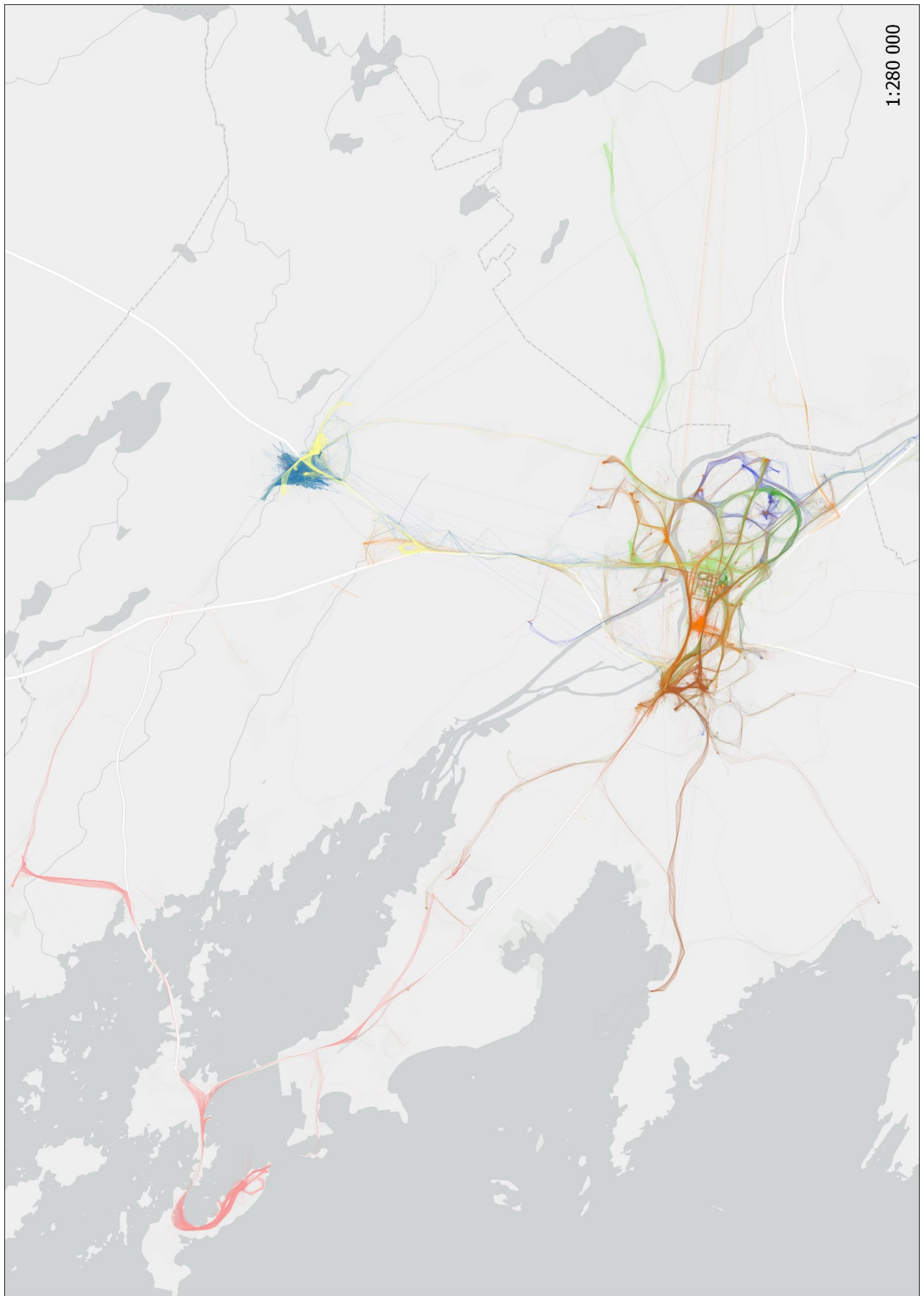
## Liite 2. Liikeratojen visualisointi



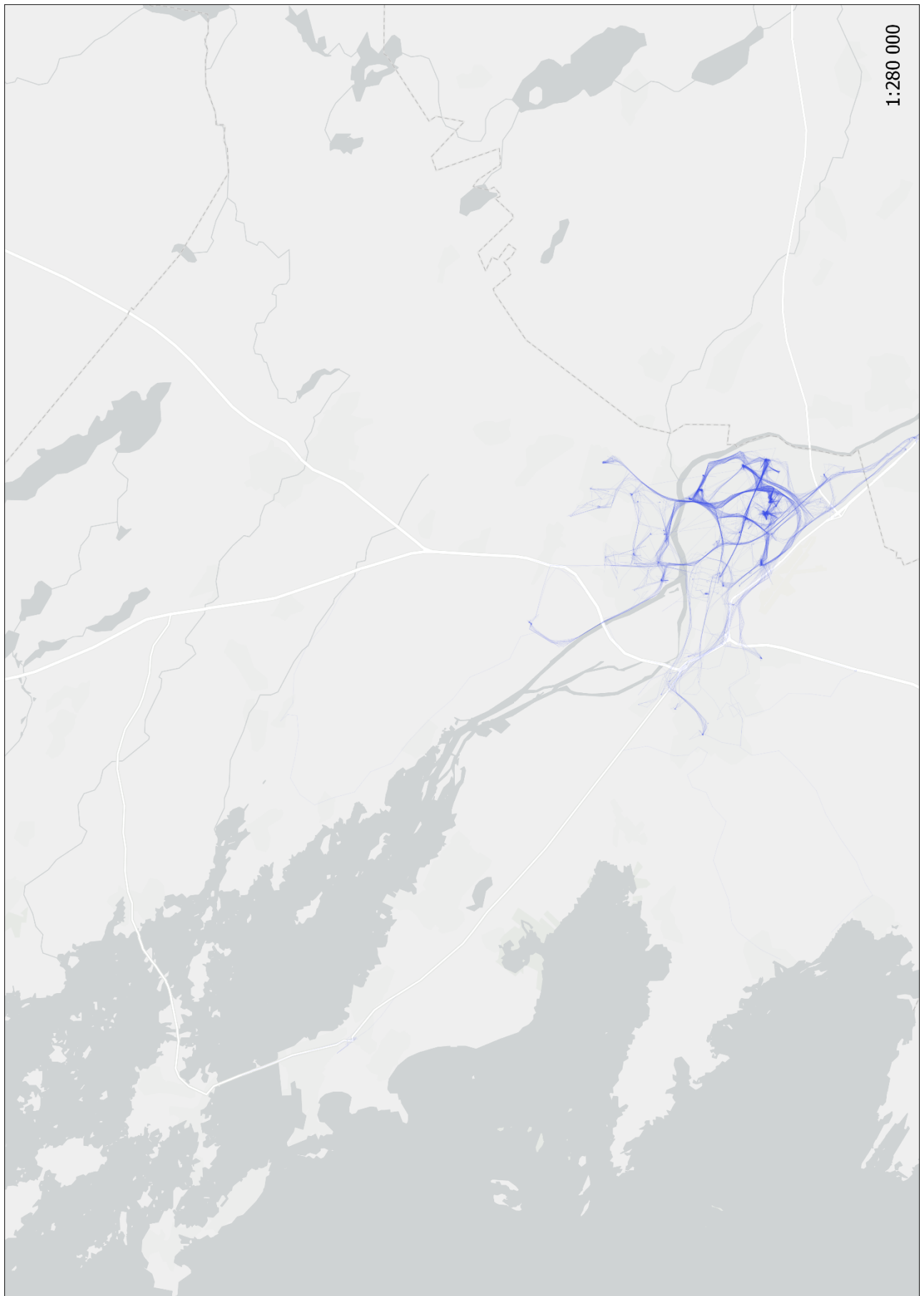
## Kaikki liikeradat

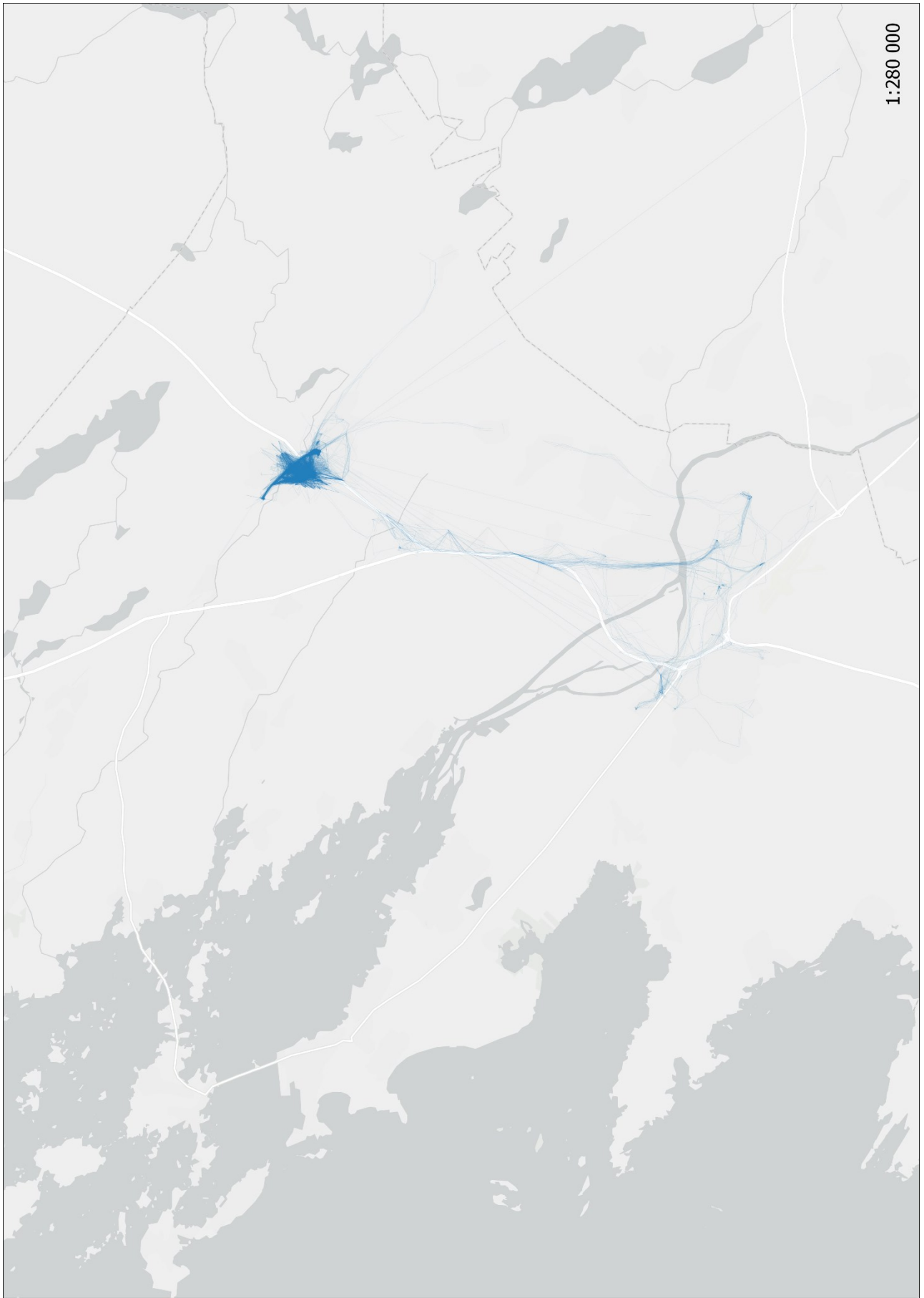


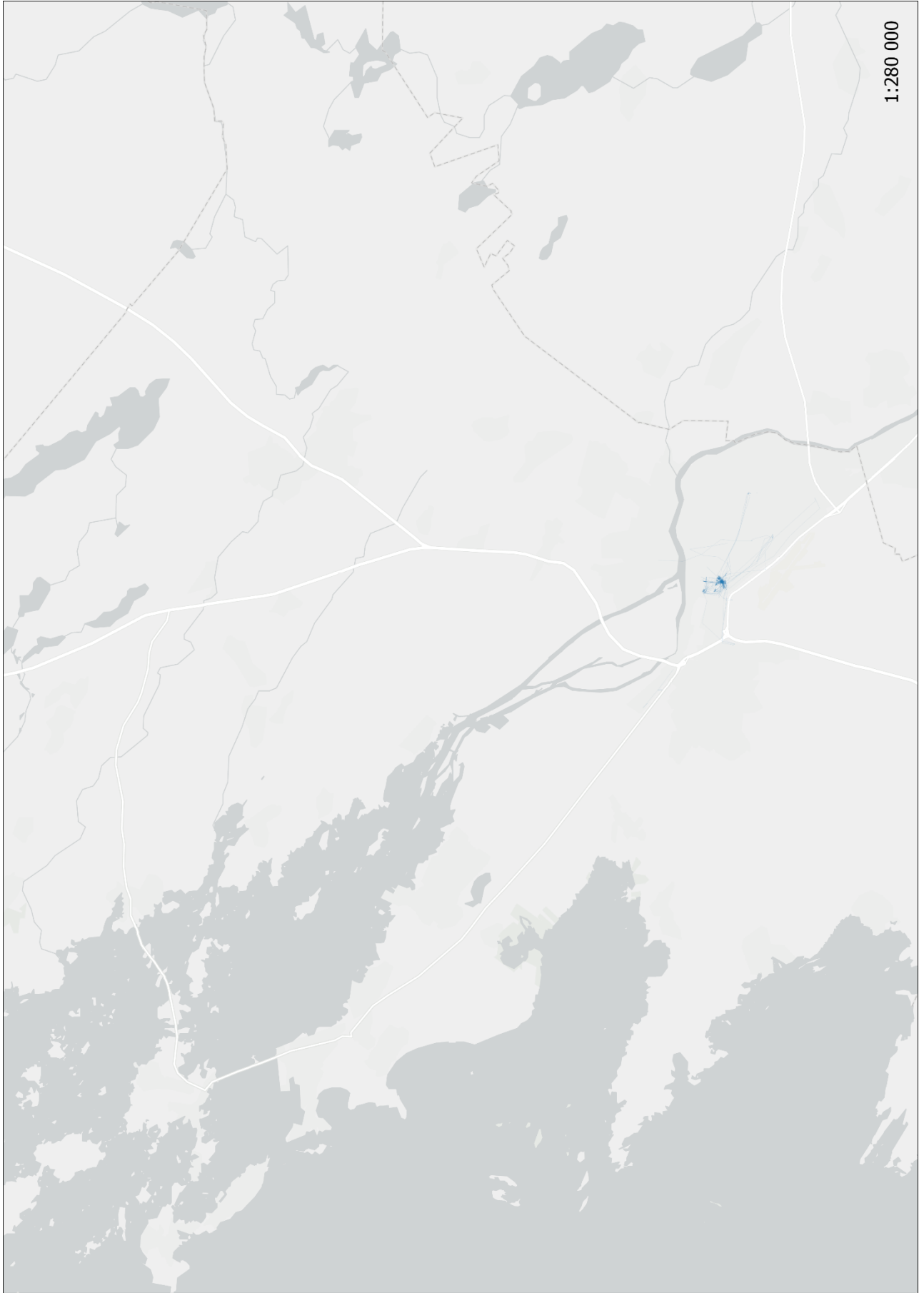


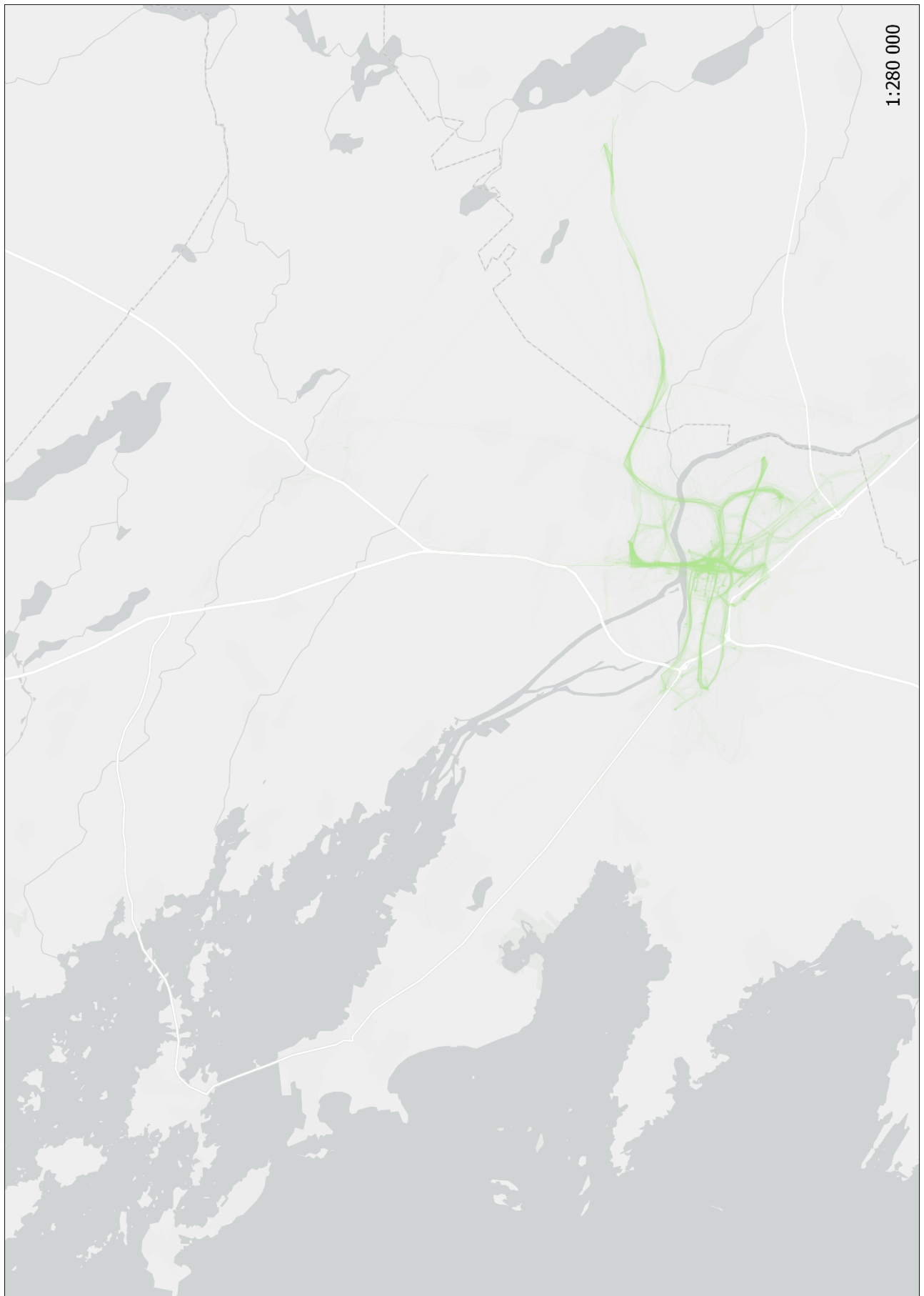


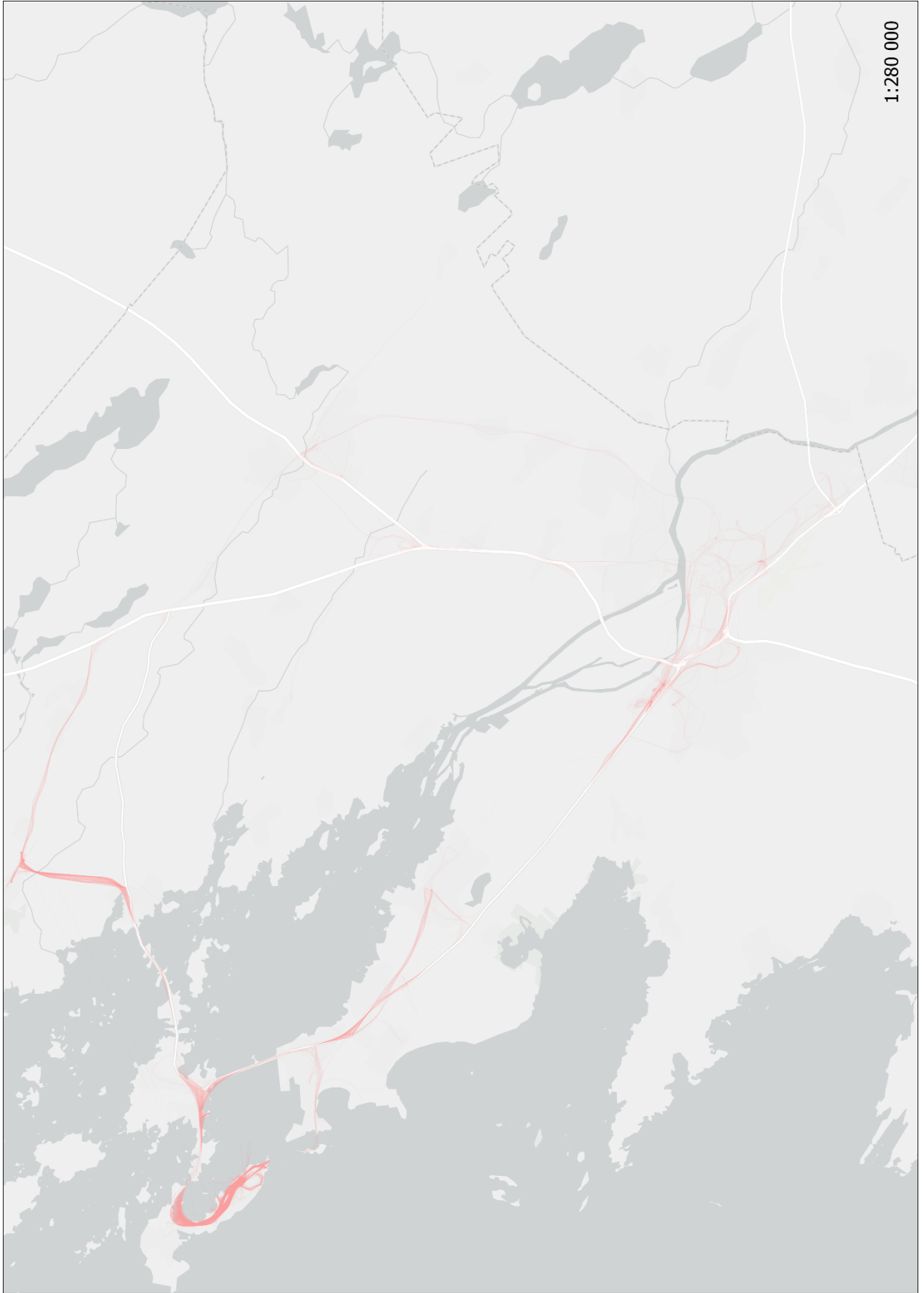




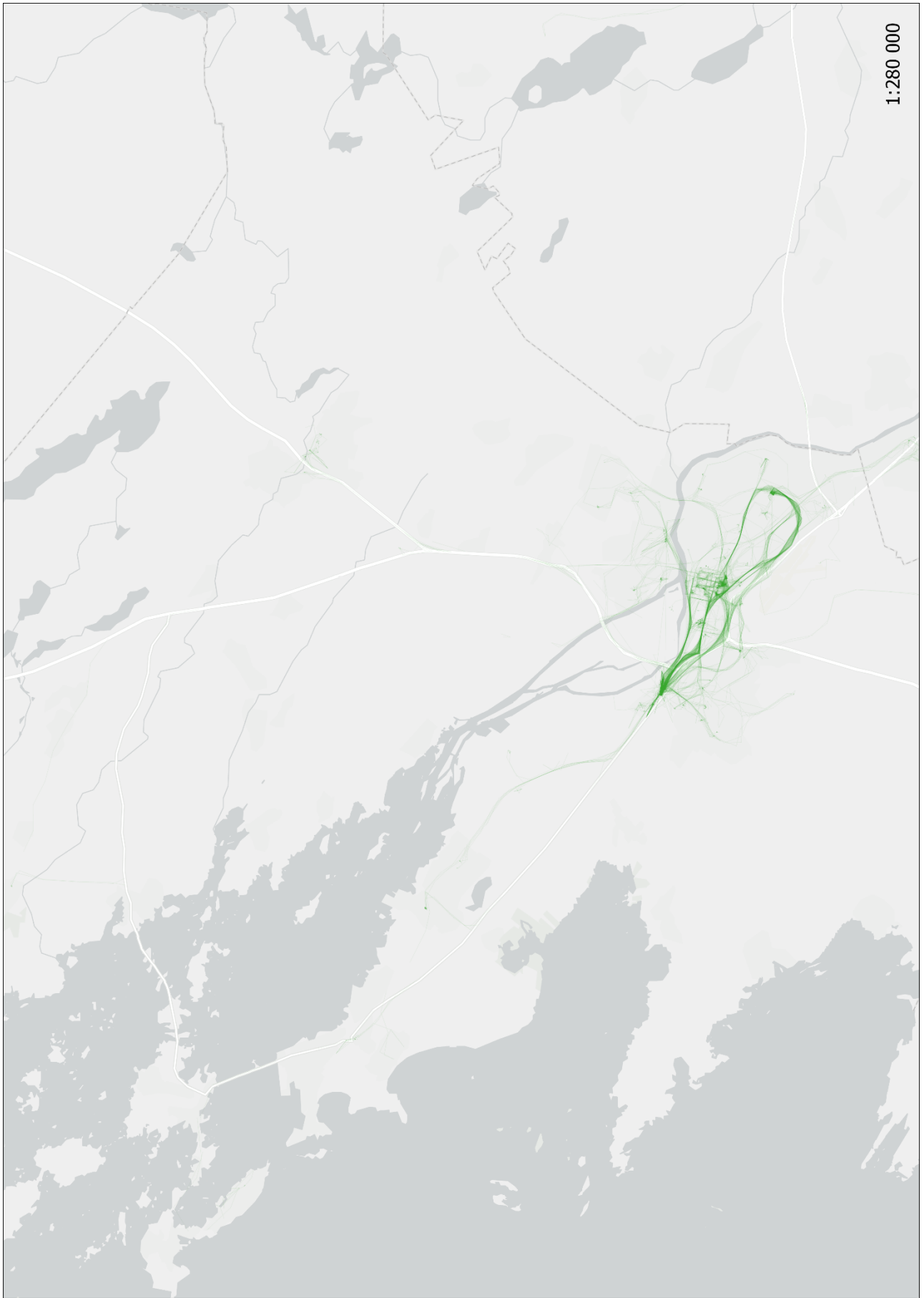


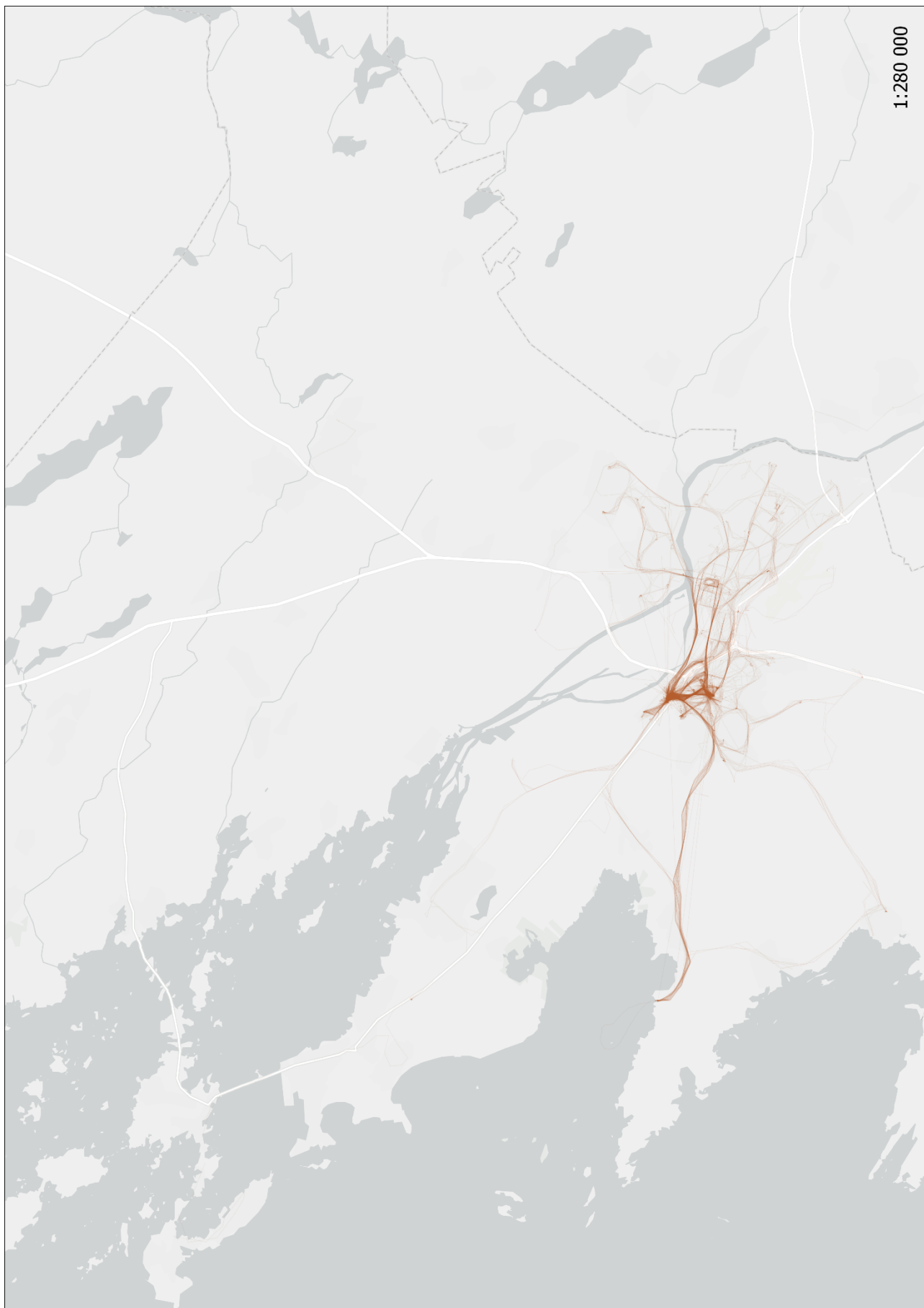




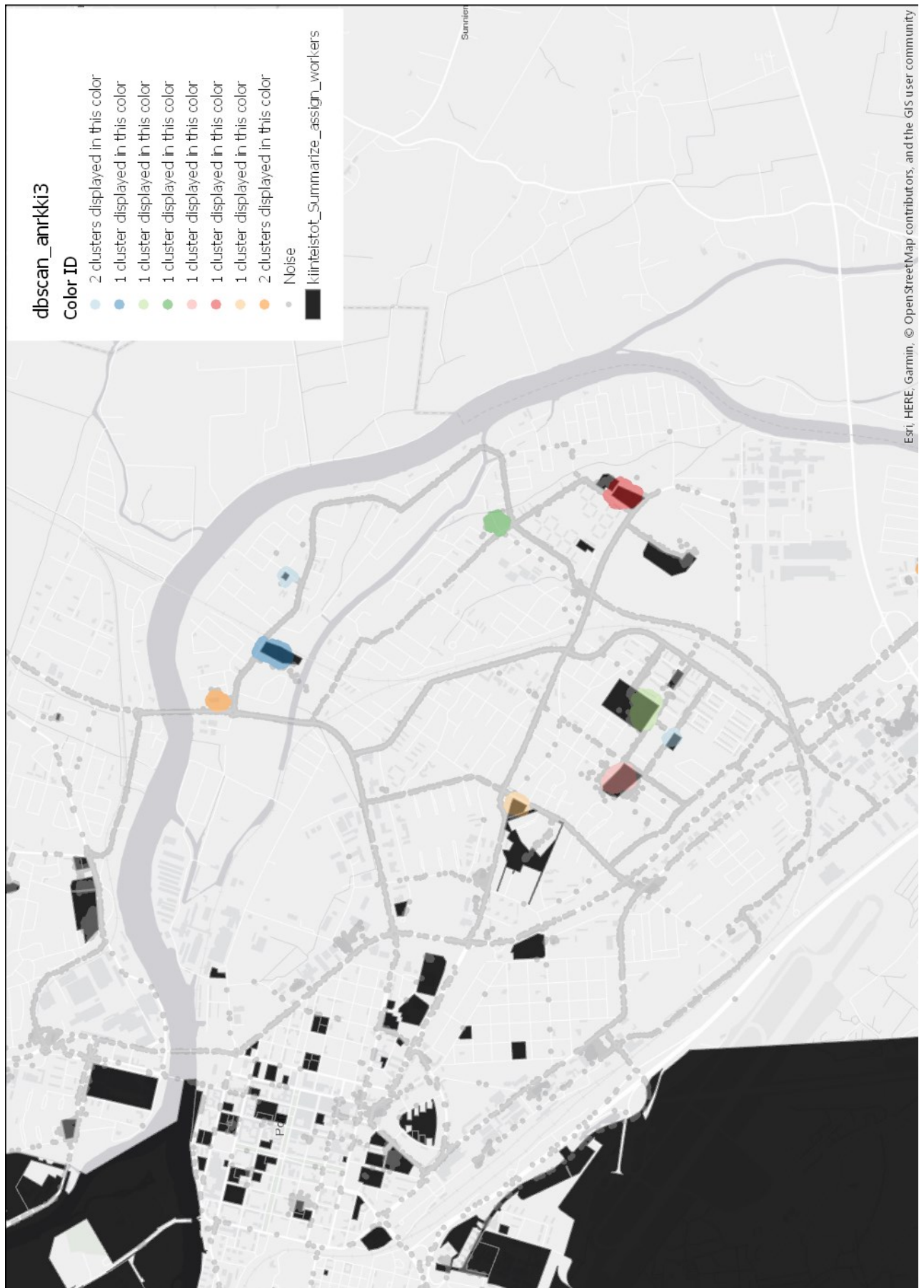


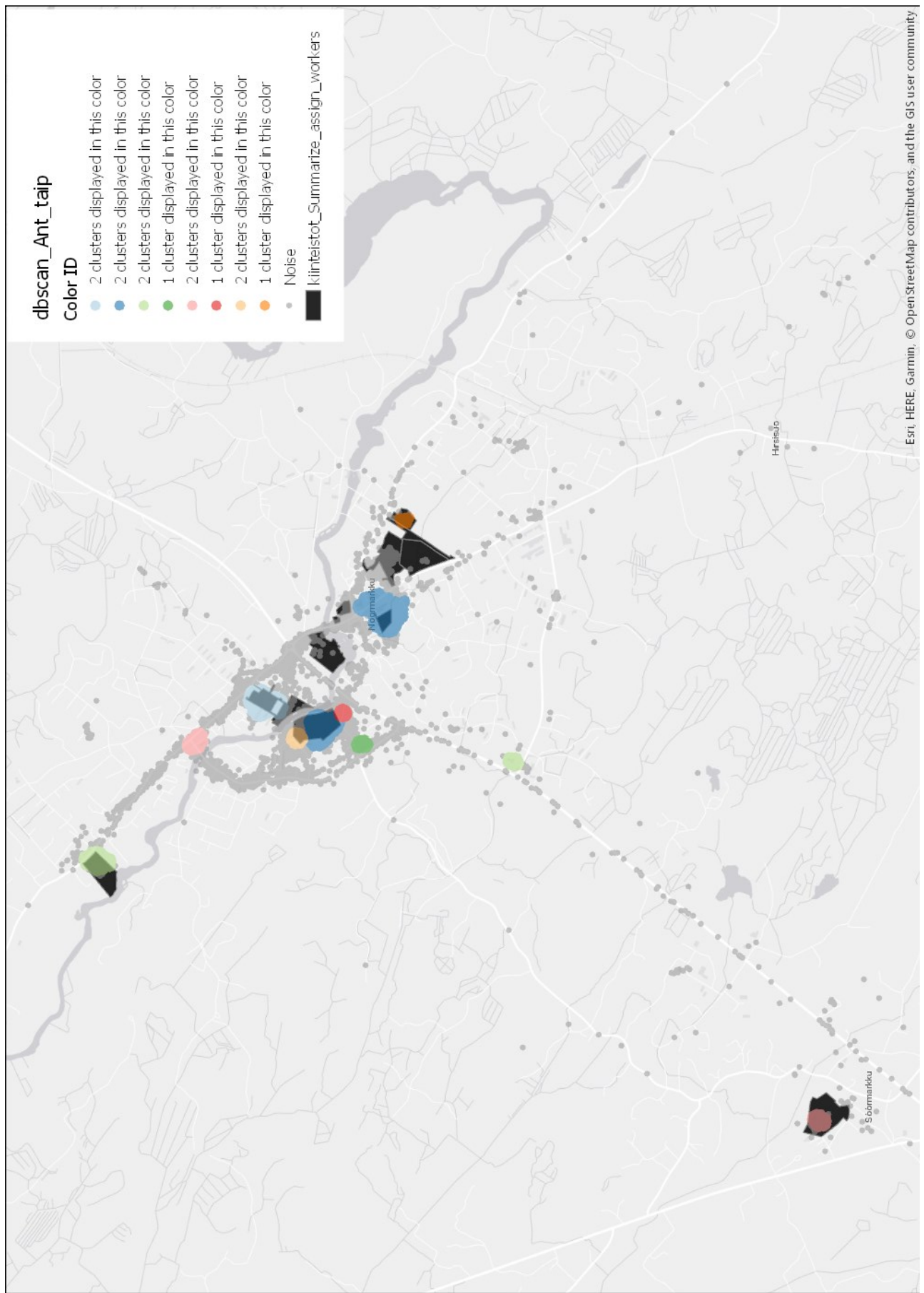




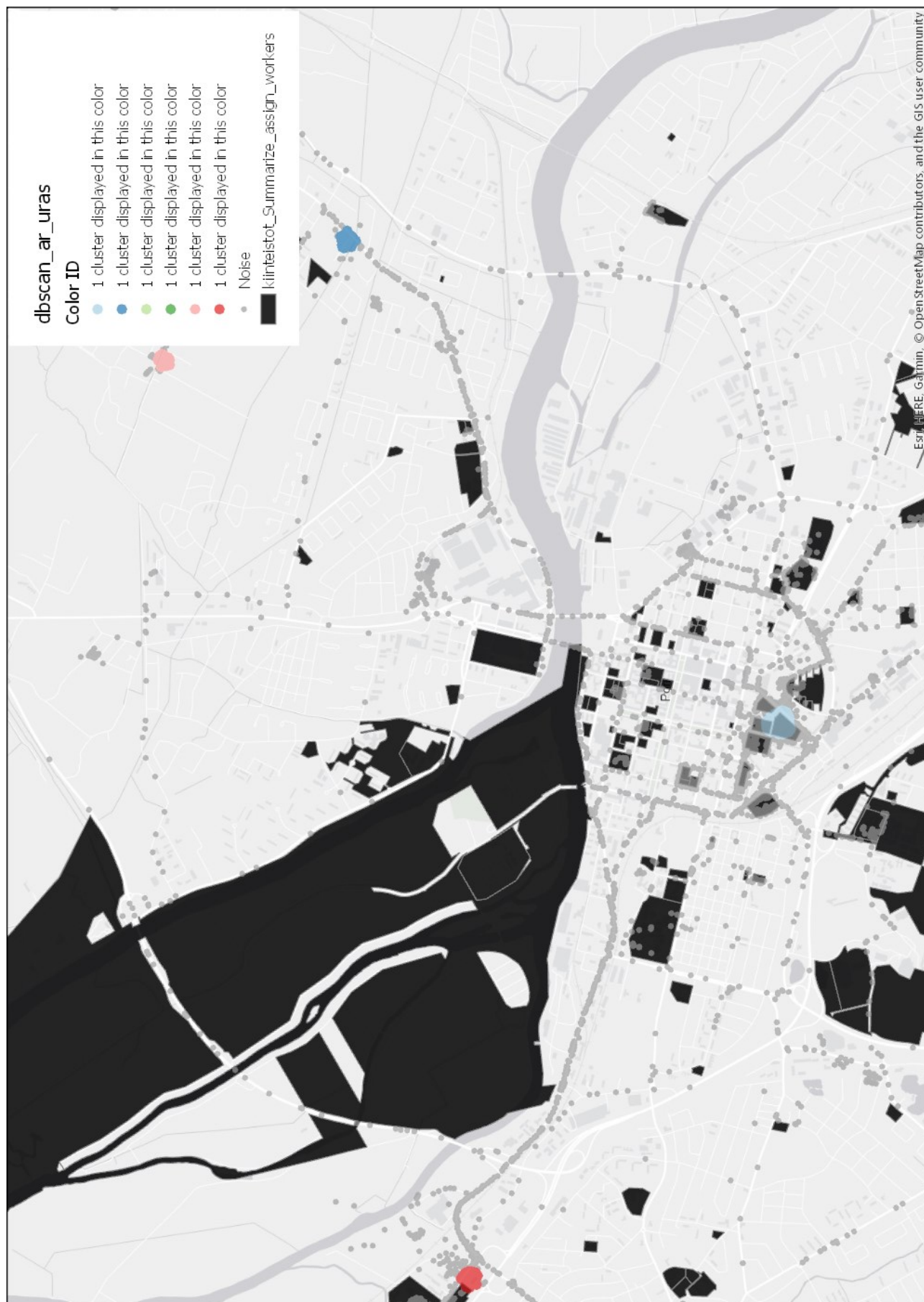


### Liite 3. Klusterianalyysi, yksittäiset työntekijät







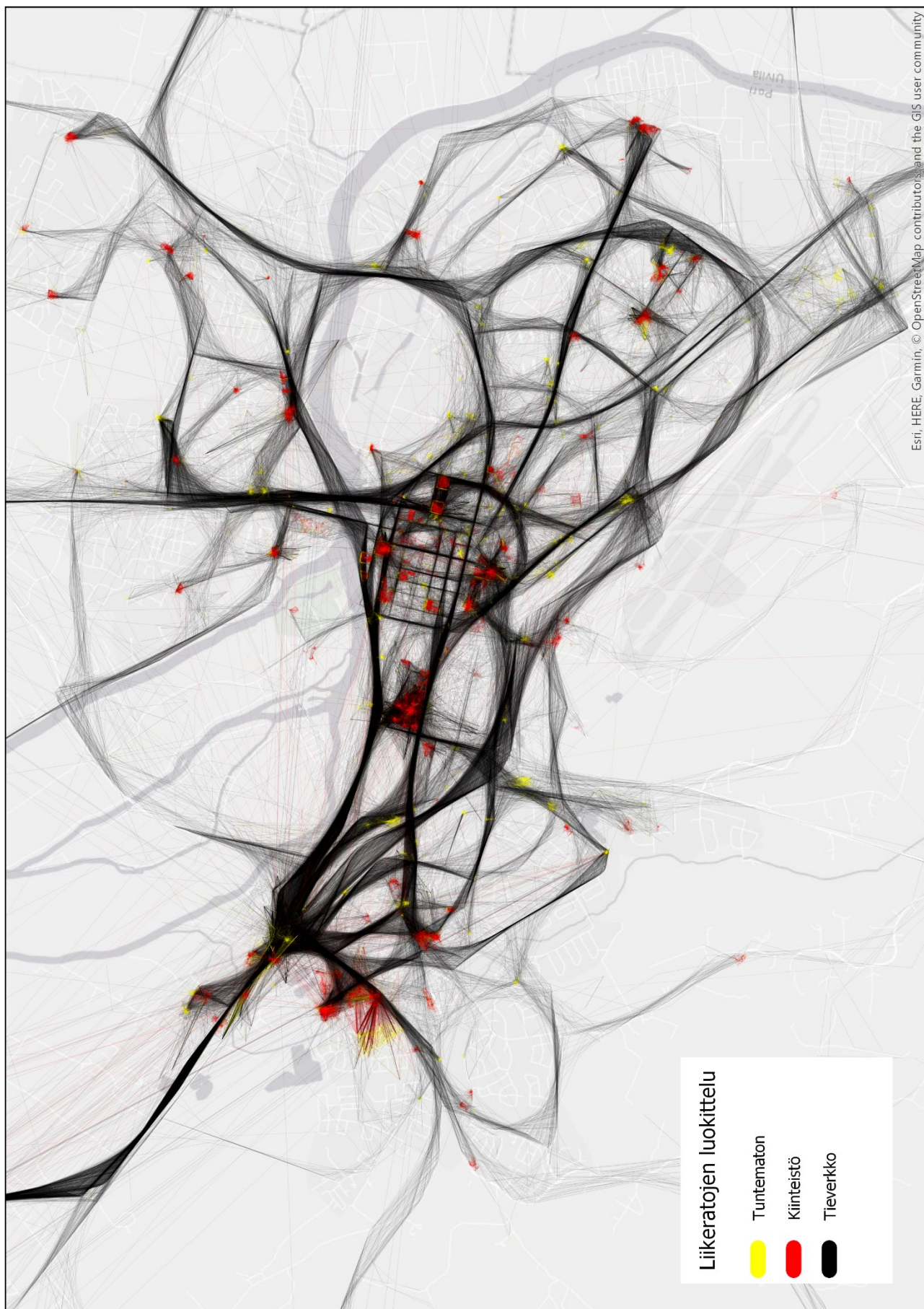


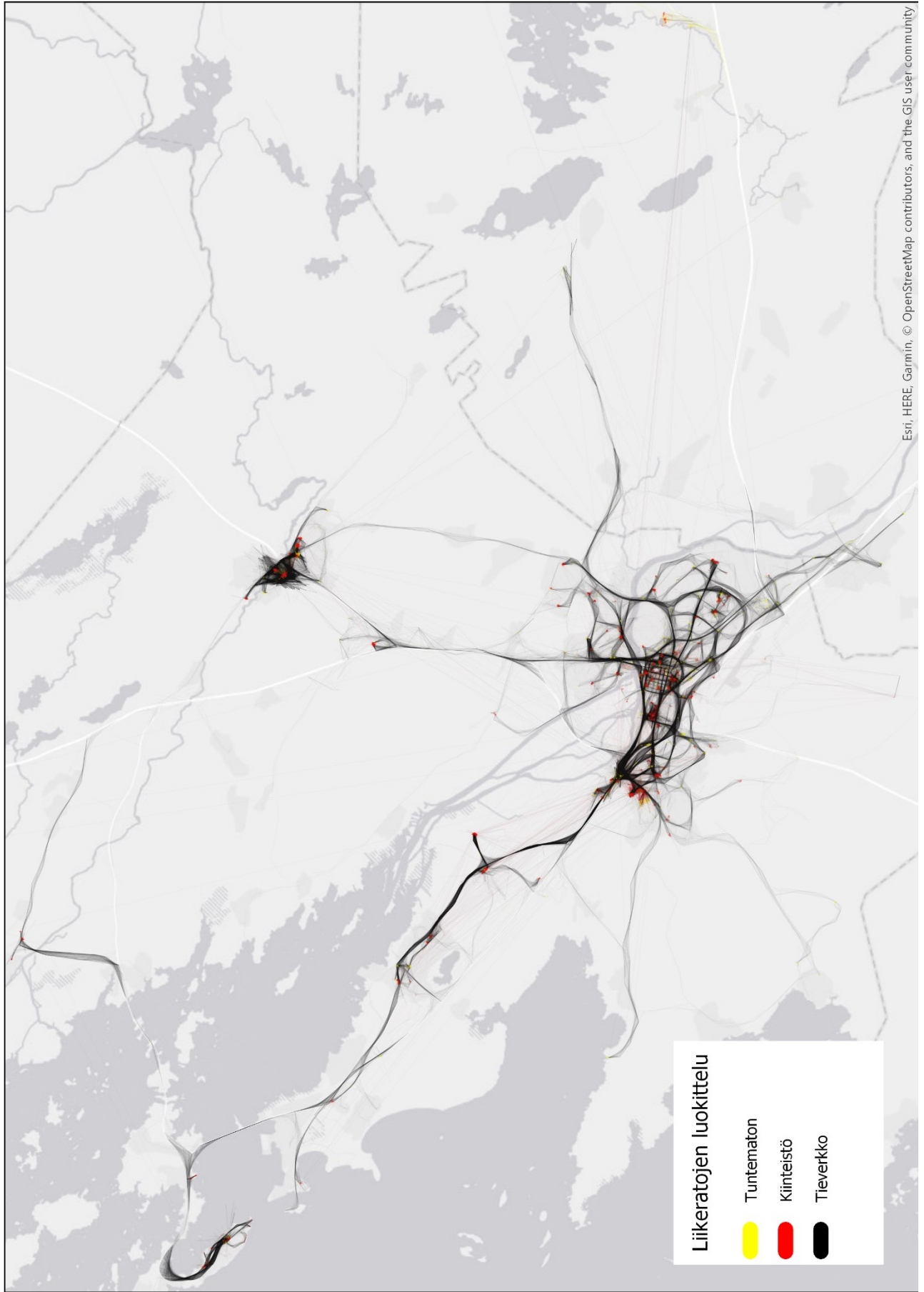


## Liite 4. Luokittelun tulokset



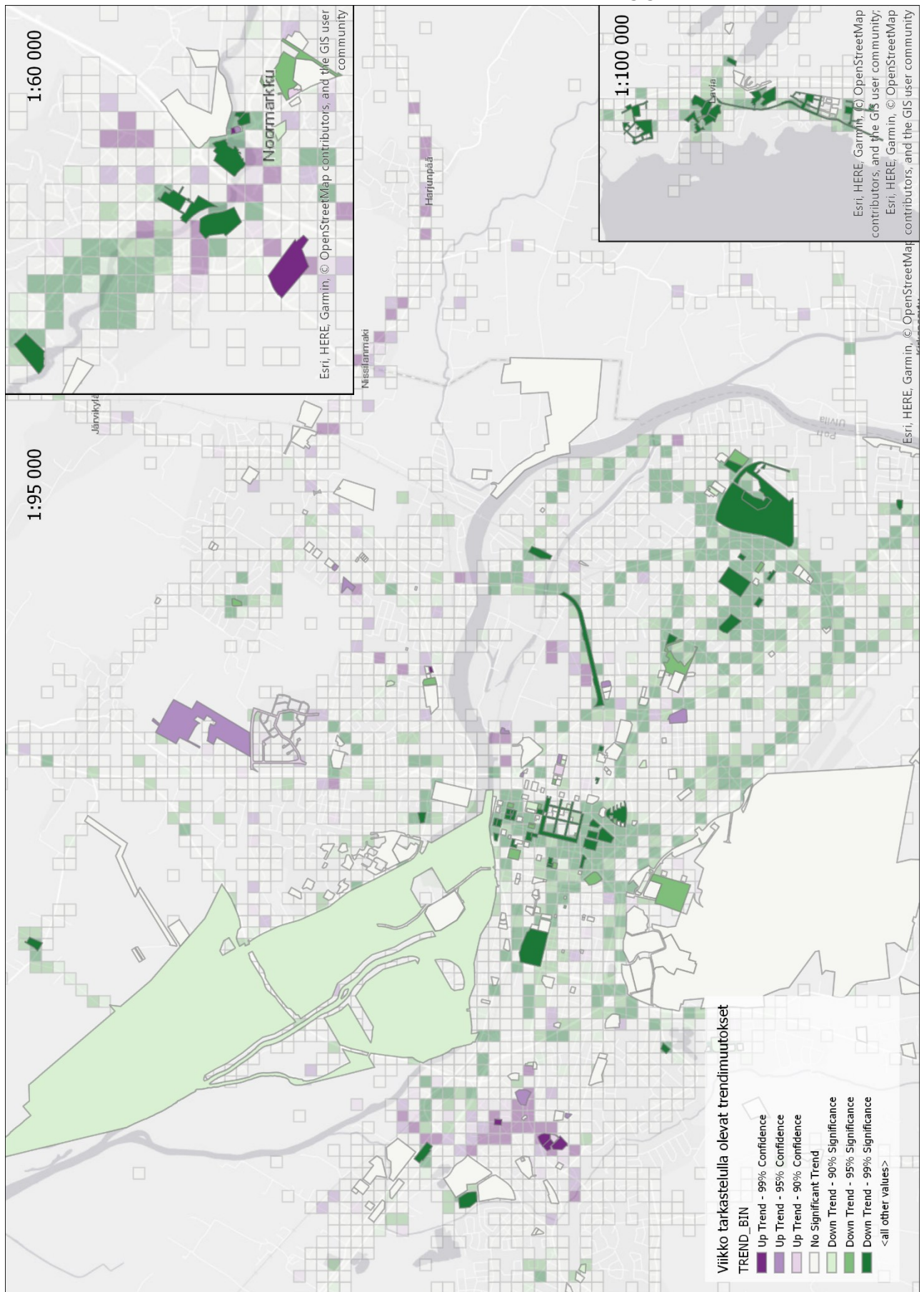


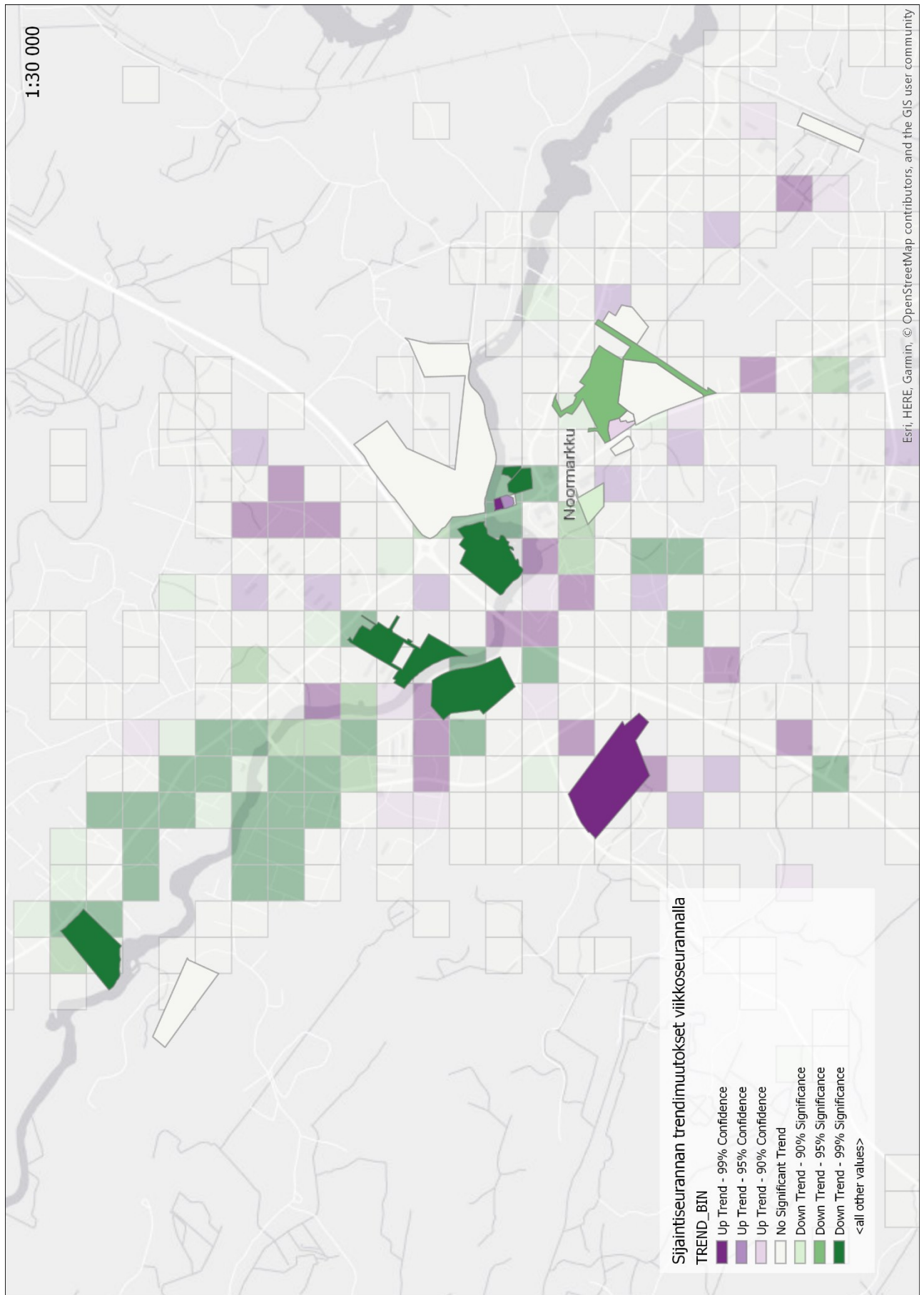






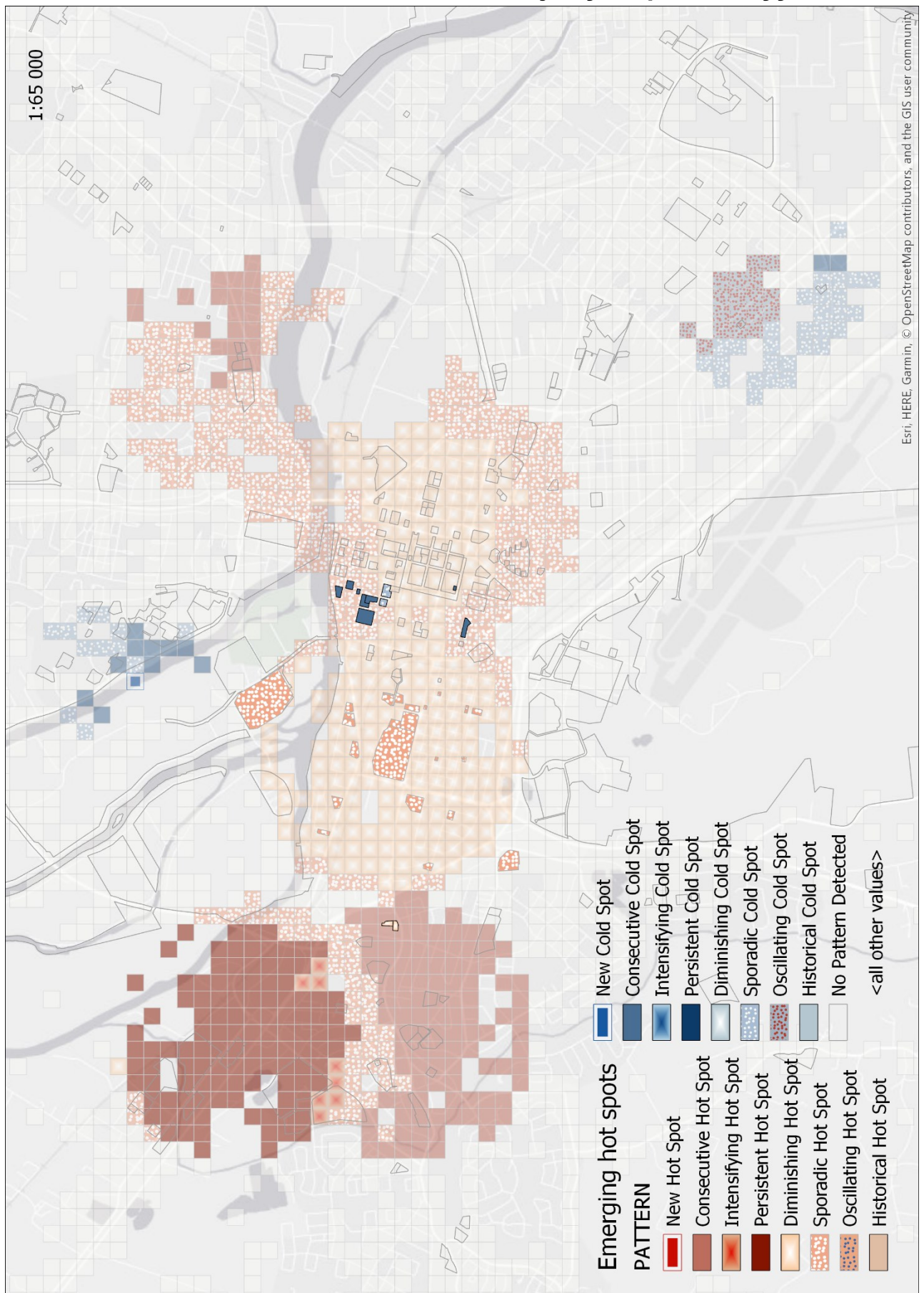
## Liite 5. Aika-avaruuskuutioiden trendianalyysitulokset

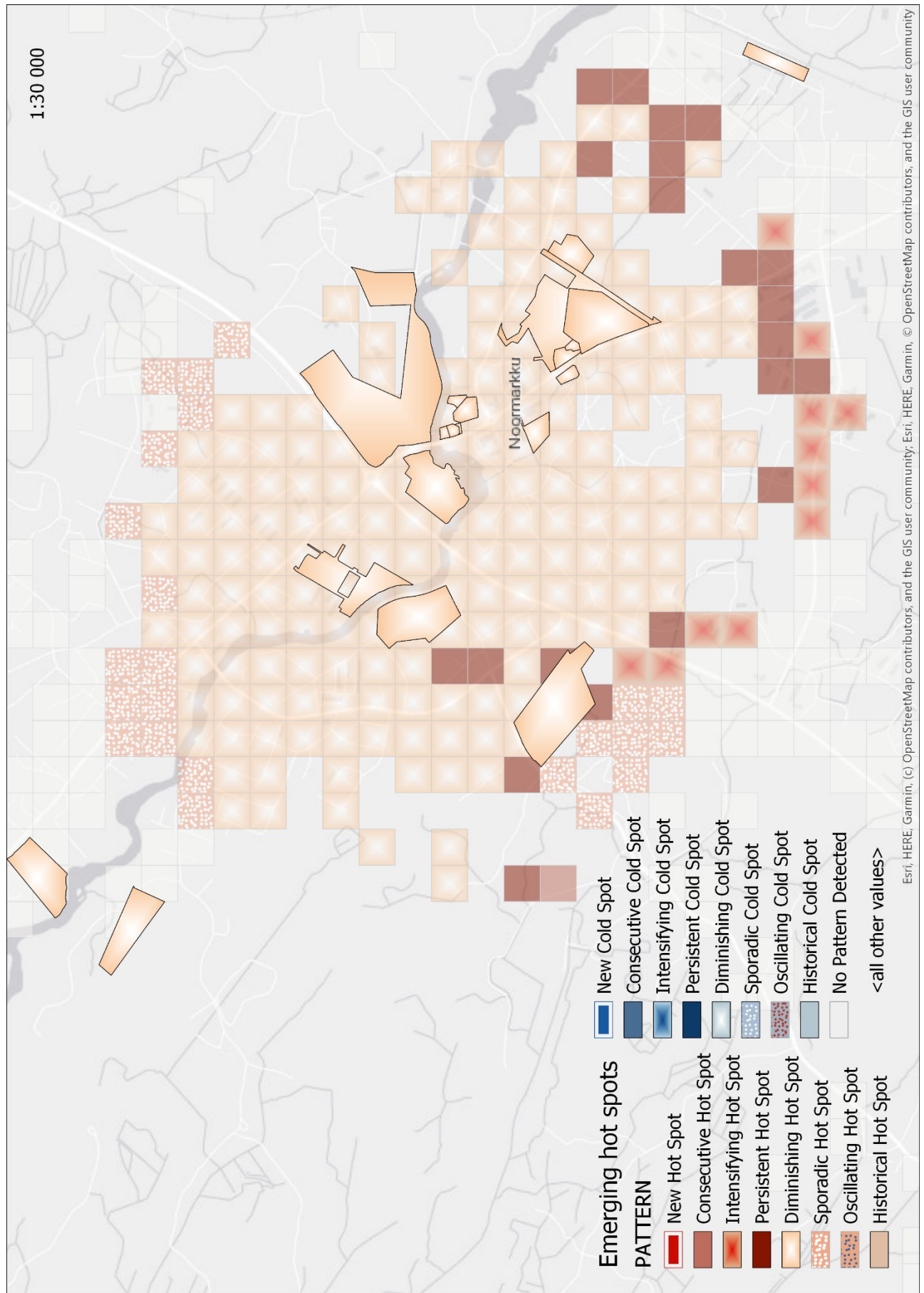




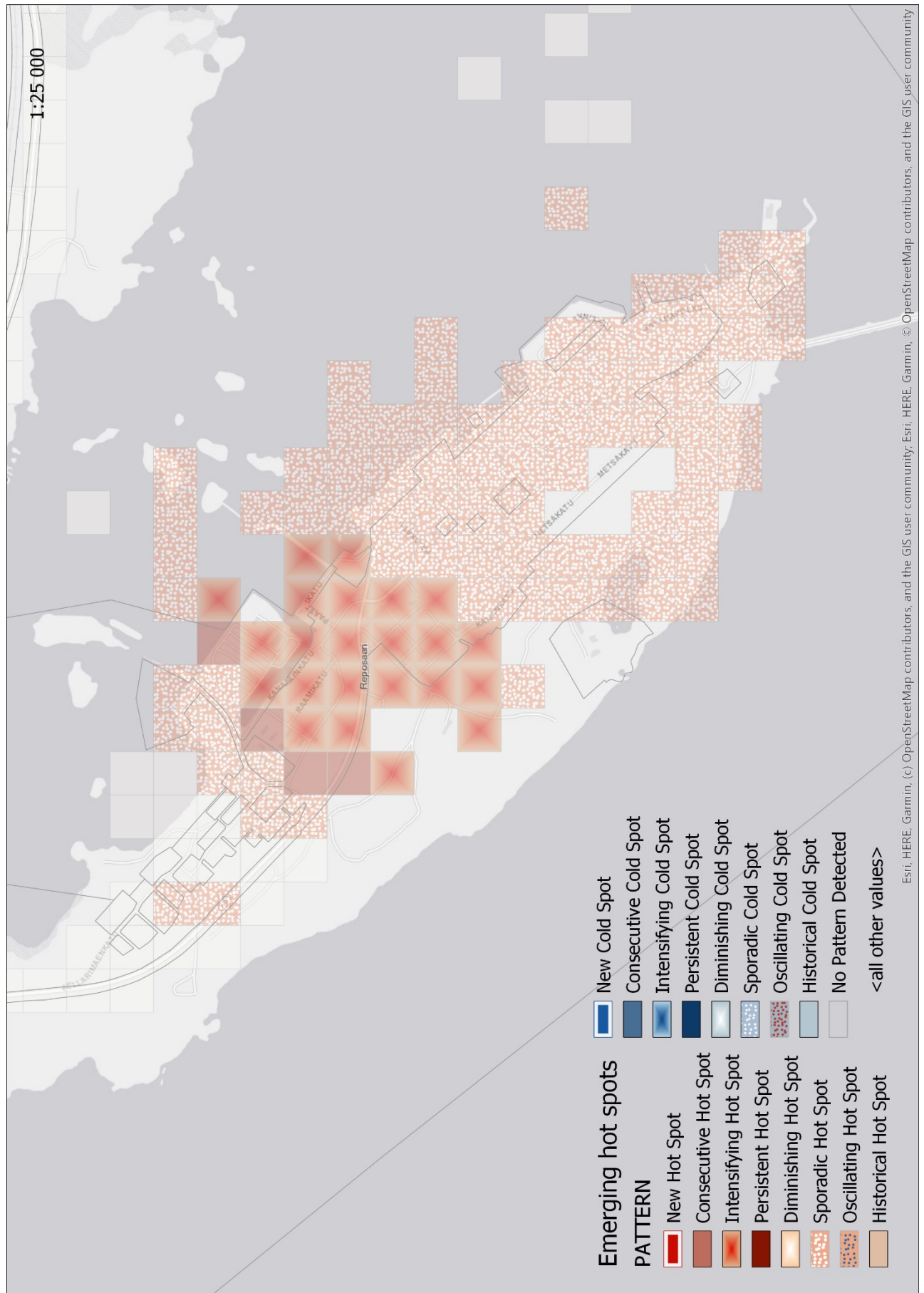


## Liite 6. Aika-avaruuskuutioiden kuuma- ja kylmäpisteanalyysit

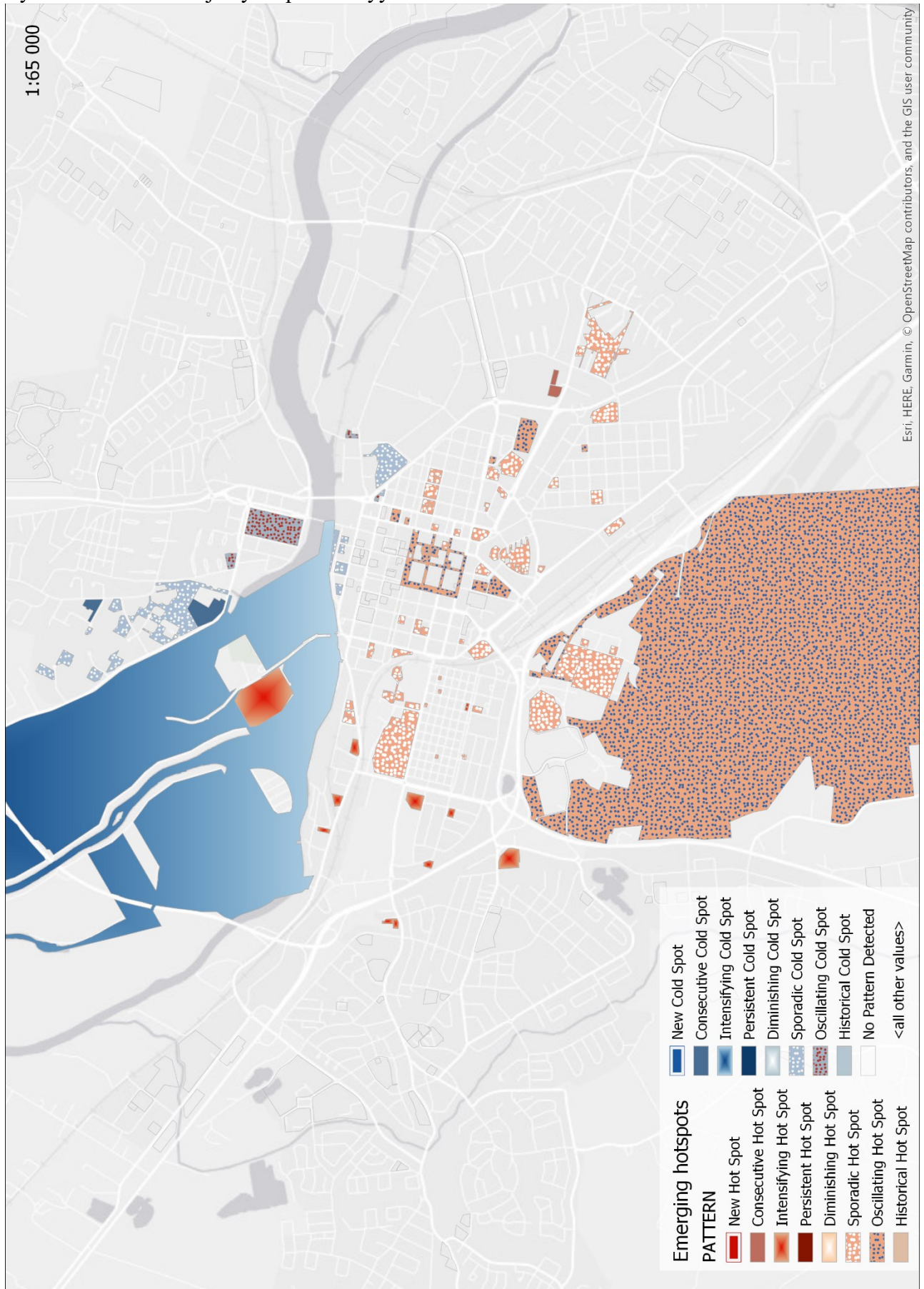




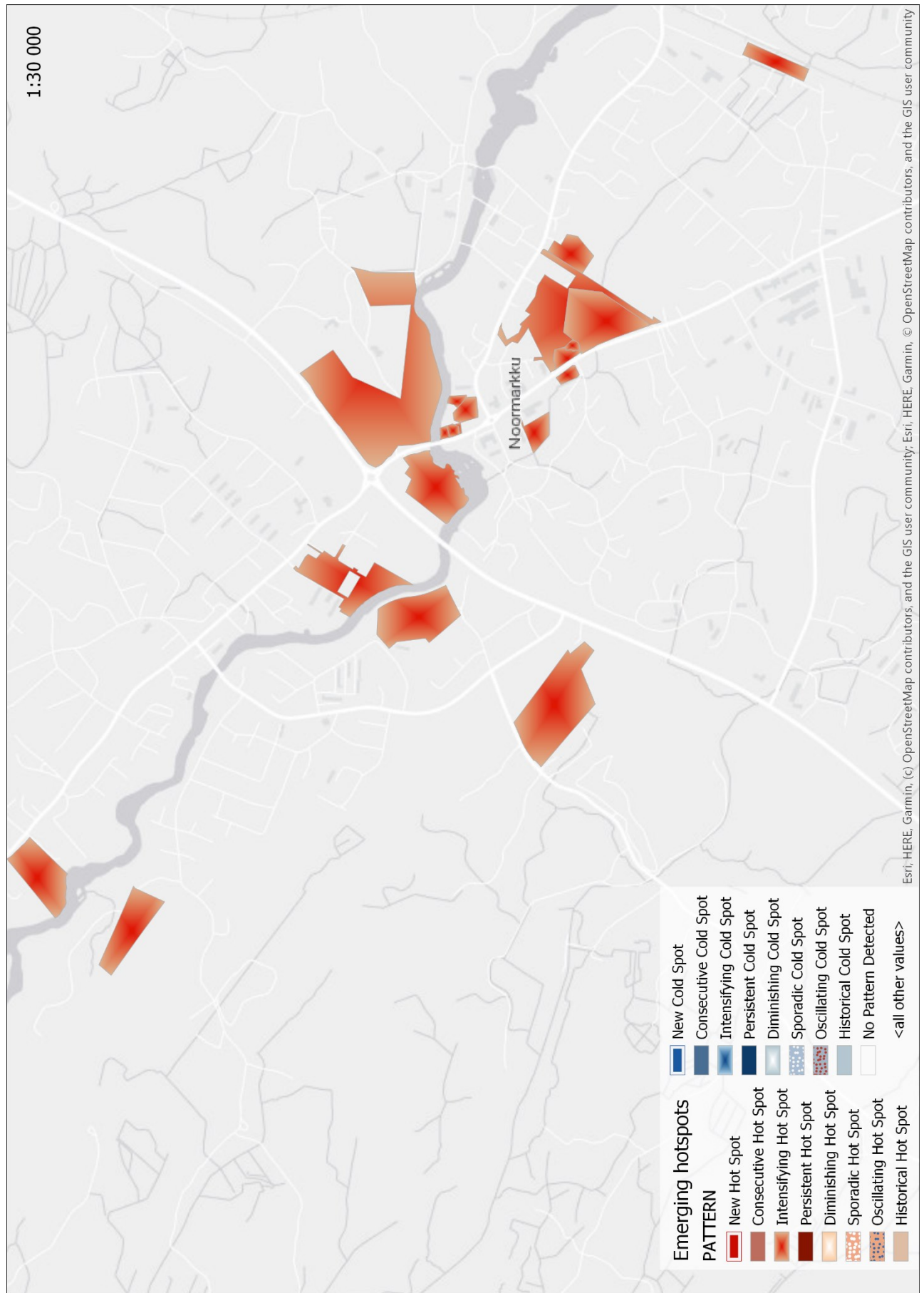




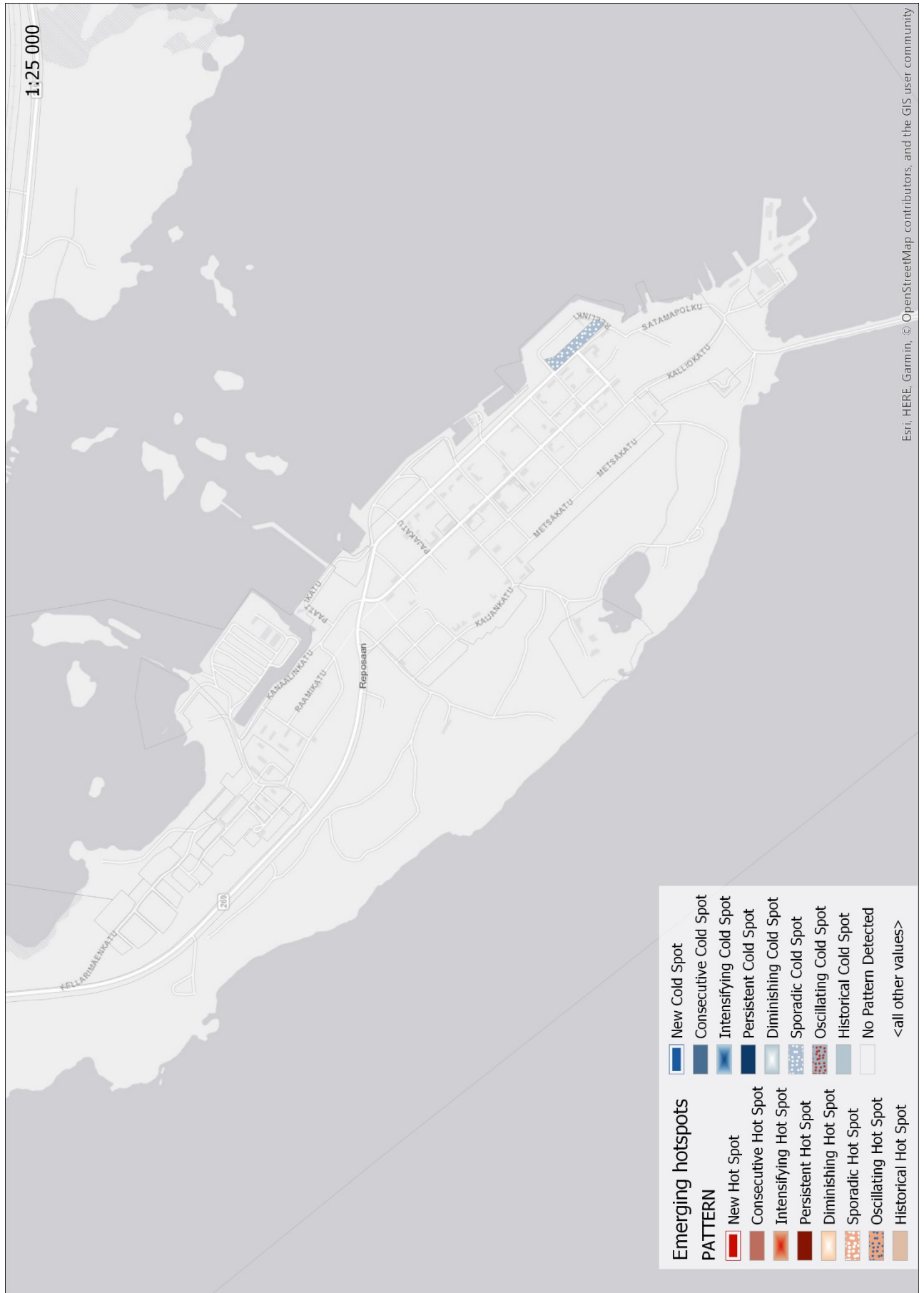
## Työtehtävien kuuma- ja kylmäpisteanalyysitulokset



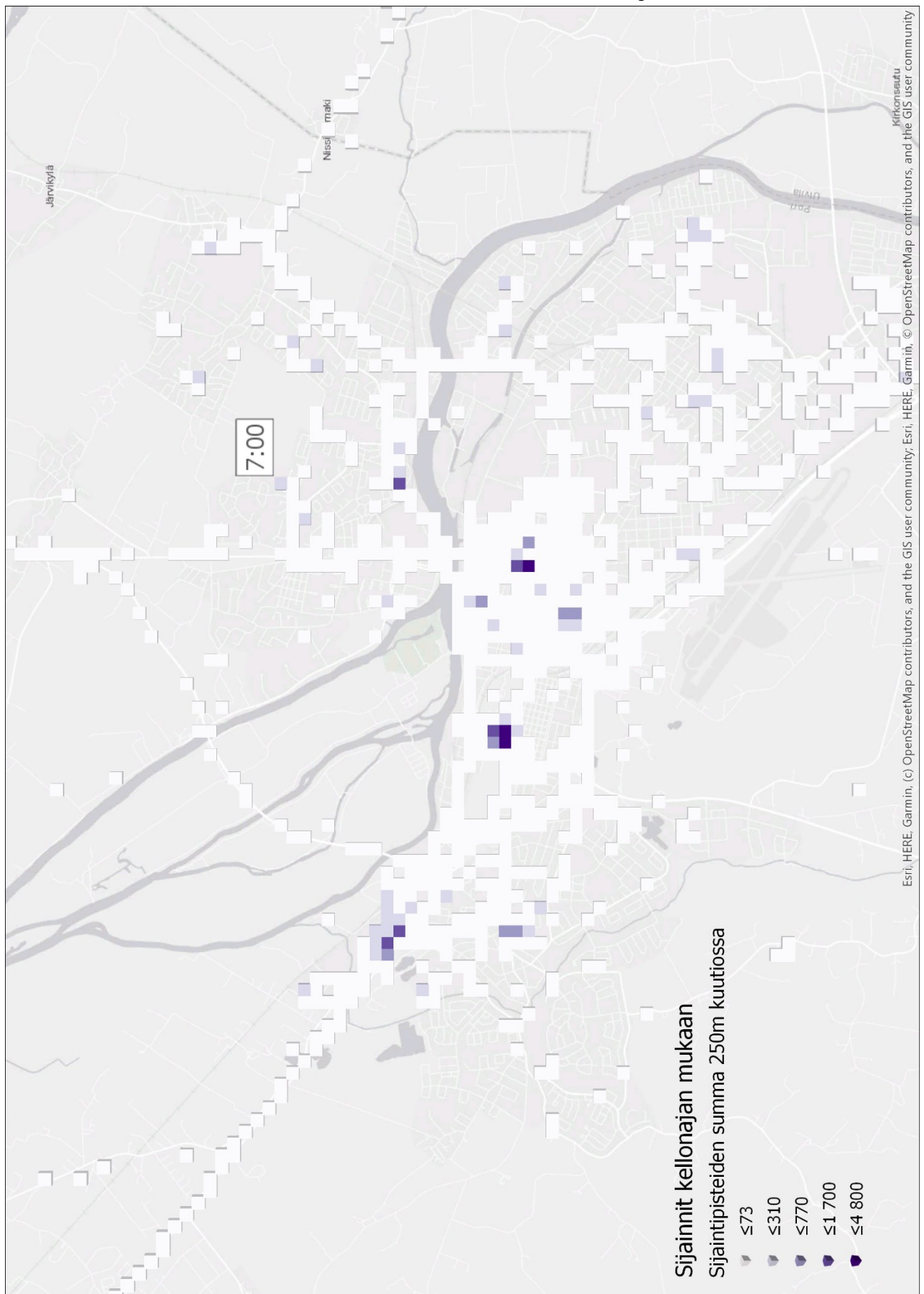


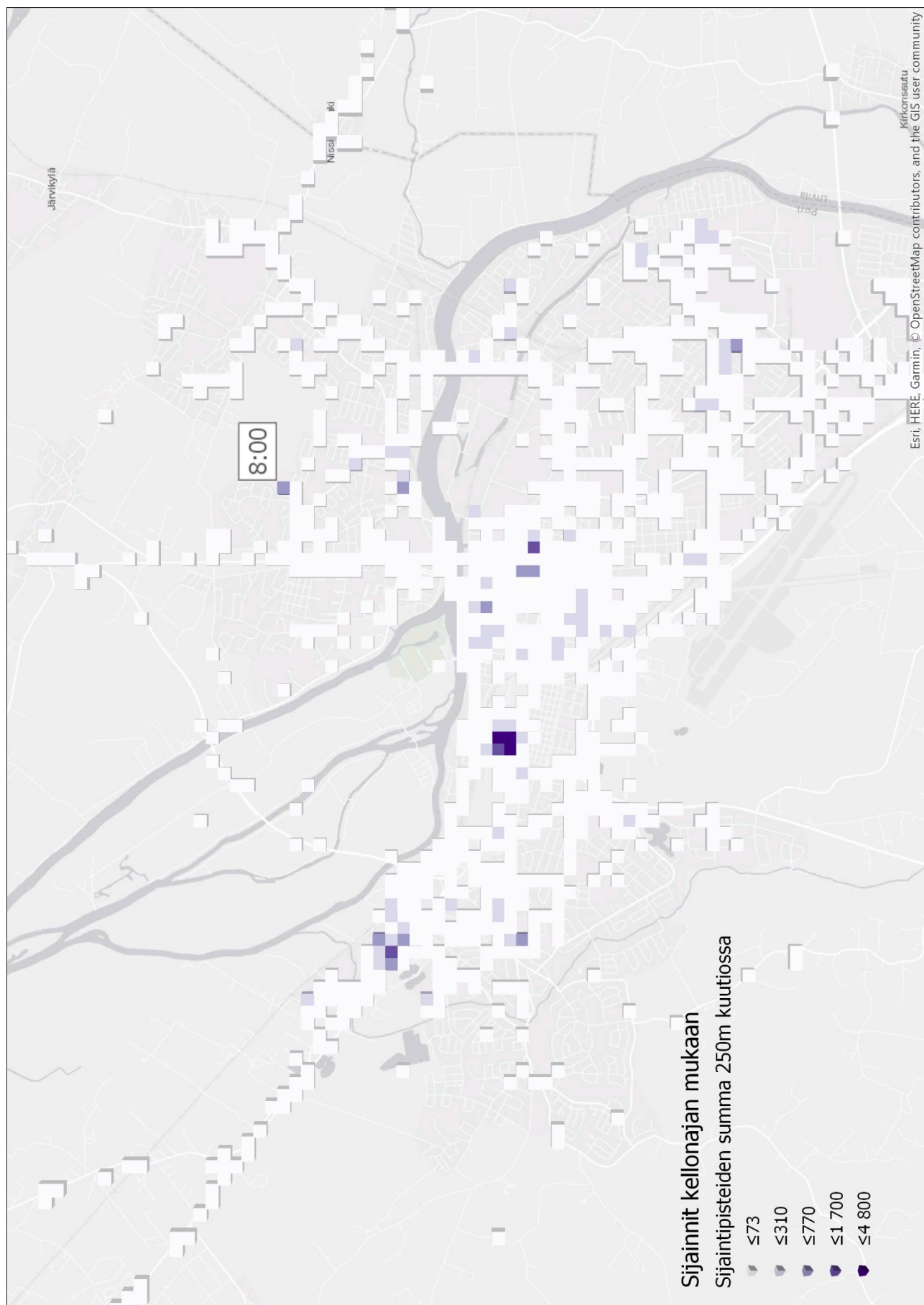


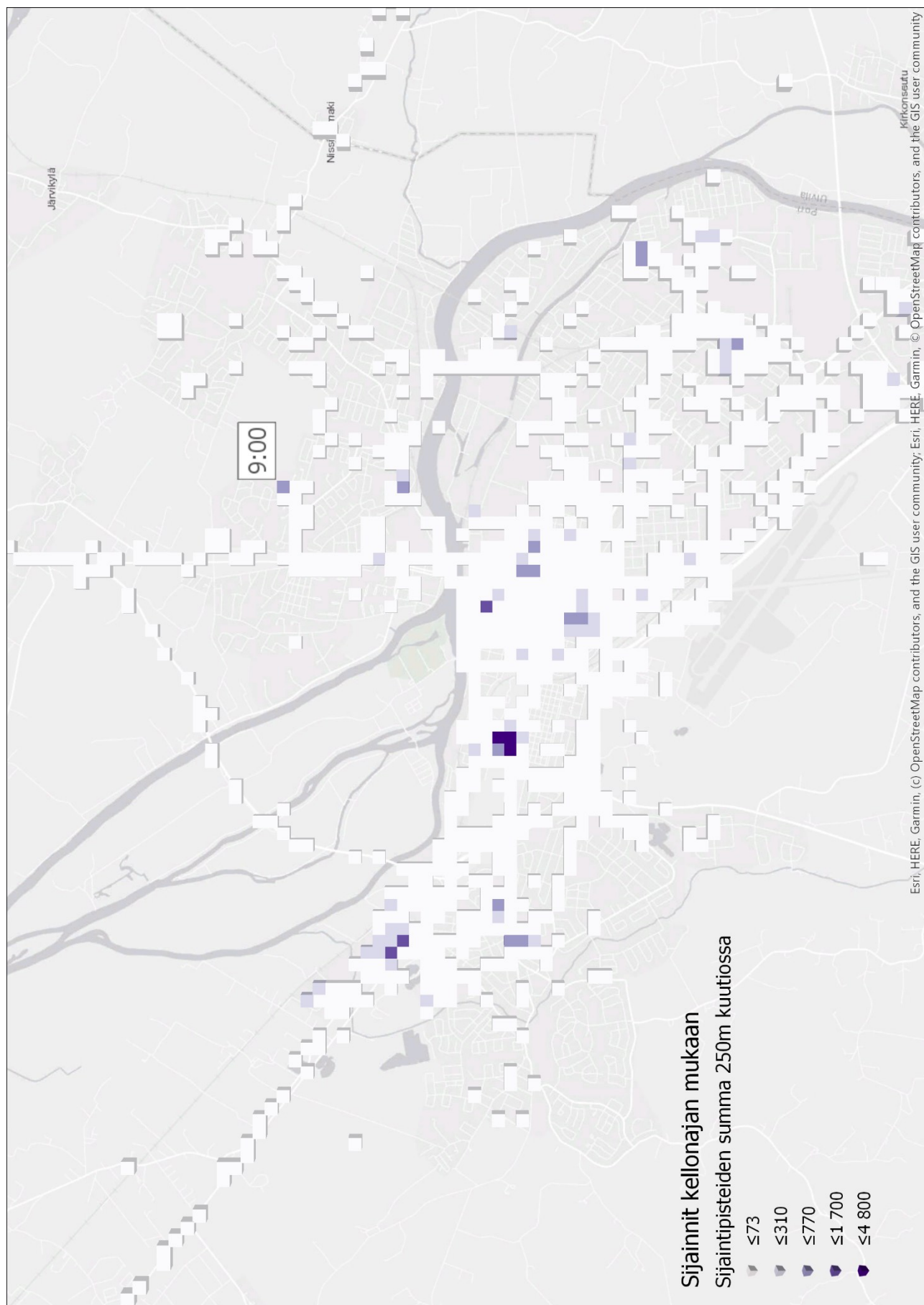




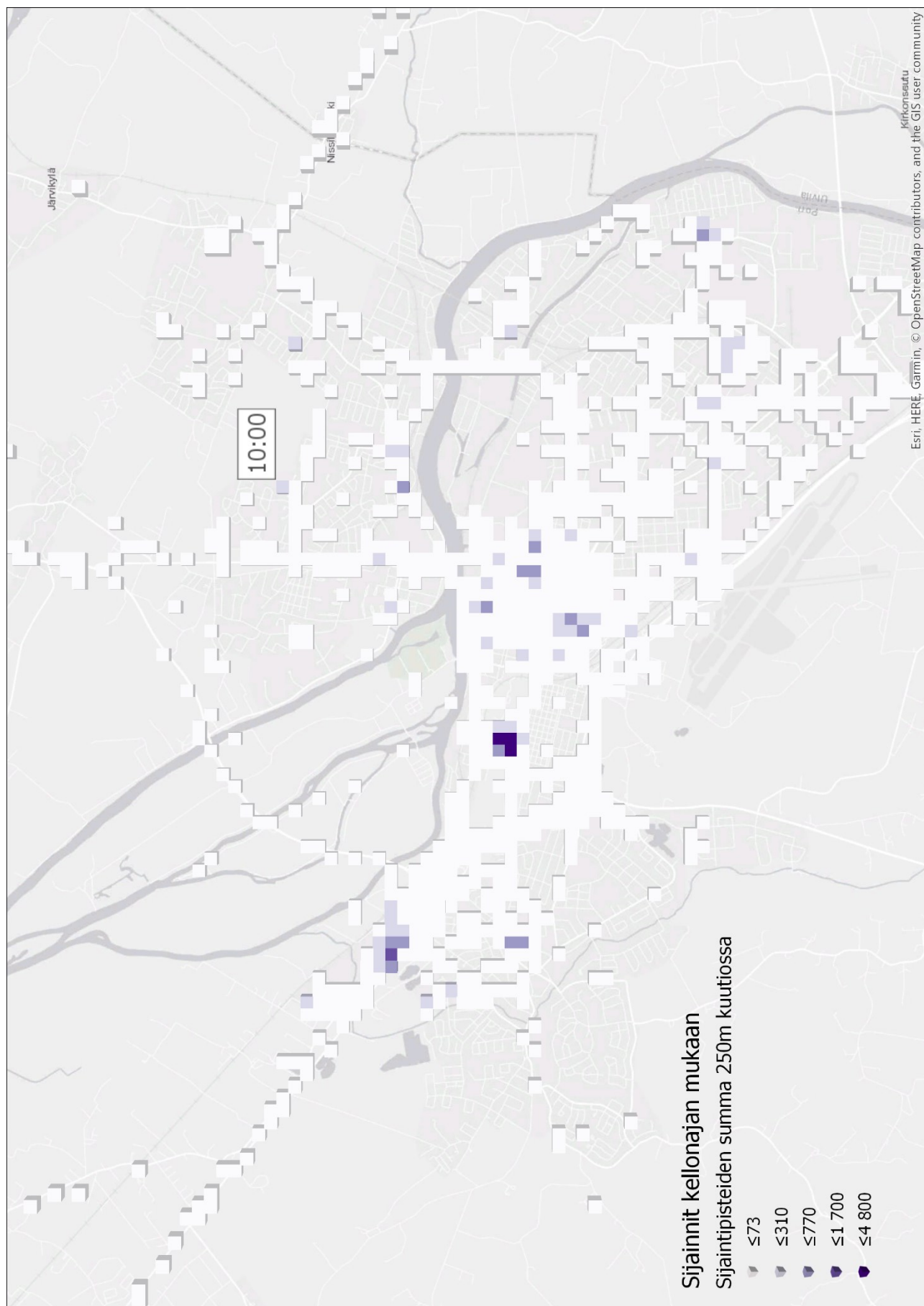
## Liite 7. Aika-avaruuskuutioiden tulokset kellonajan mukaan



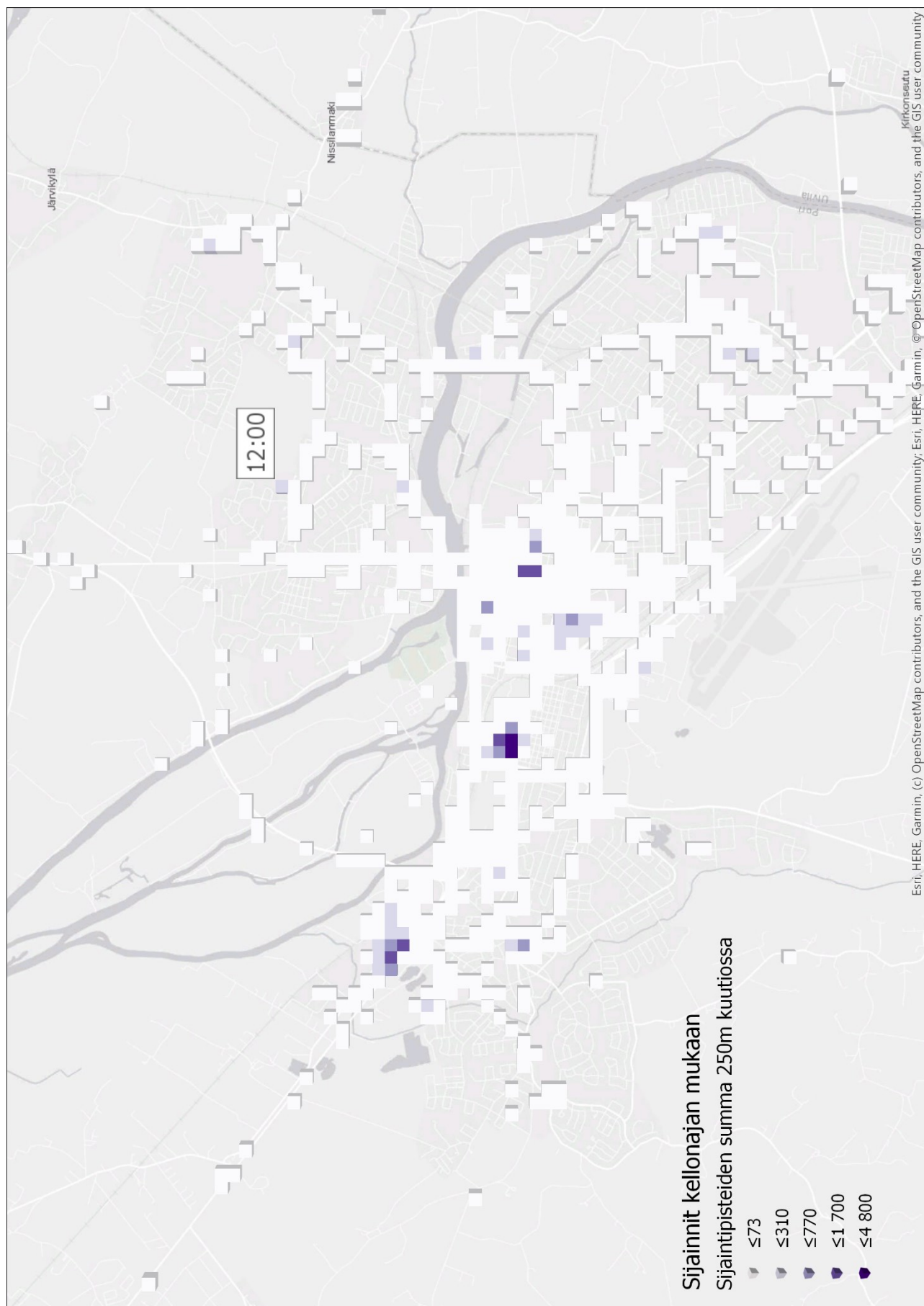


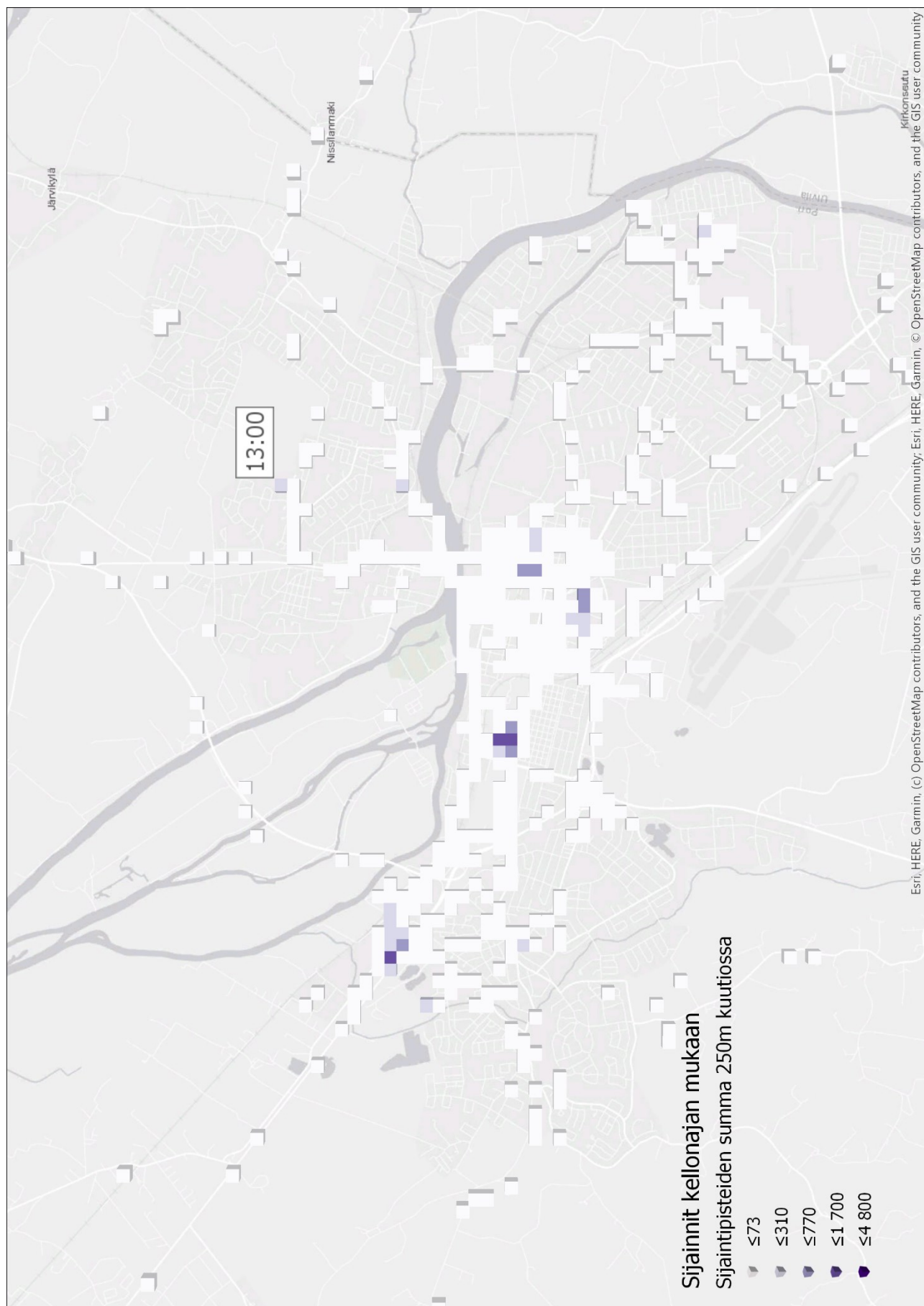


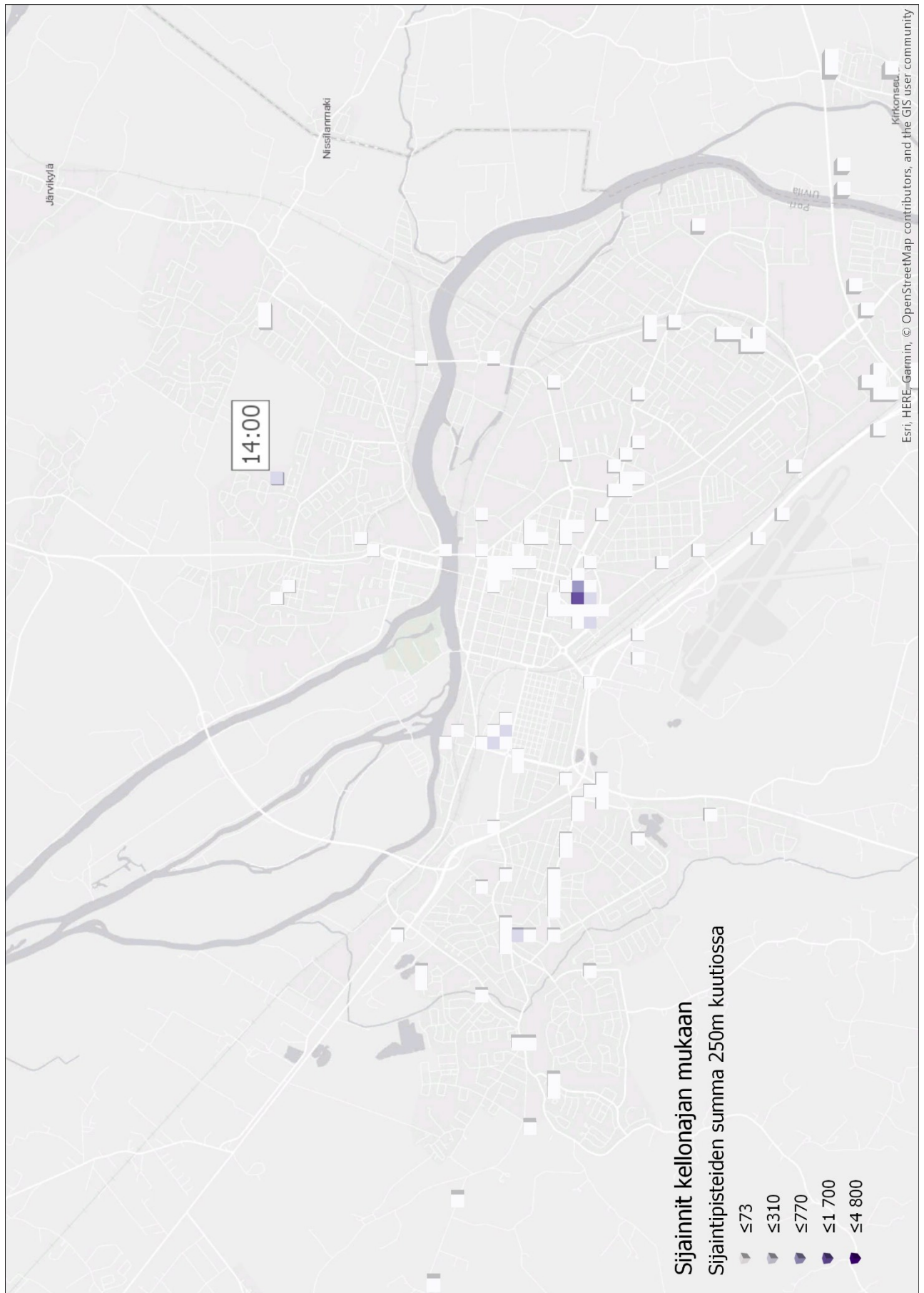












## **Liite 8. Porin kaupungin kommentit työn tuloksista**

Kommenttoija DI Mikko Viitala, tilayksikön päällikkö.

### **Automaattinen tekstiviestijärjestelmä**

“On hyvä, että tähän saatiin ratkaisu. Koska palvelupyynnot lähettää usein kaupungin sisäinen toimija, niin molemmat tavat sopii sekä sähköposti tai tekstiviesti. Tämän toteutusmalli pitää testata nyt käytännössä. Uskoisin, että viestintä parantaa asiakastytyvääisyyttä.”

### **Klusterointimenetelmät**

“Tälle analyysille tulevia käyttötarkoituksia on muun muassa uusien tukikohtien suunnittelu, taukopaikkojen paljastuminen, paljon työtä vaatineet kohteet ja lisäresursoinnin suunnittelu ja paljon työtä väsyneiden kohteiden tarkempi analyysi voisi johtaa järjestelmien tai koko talon peruskorjaukseen.”

### **Luokittelumenetelmät**

“Oli erittäin mielenkiintoista nähdä, miten kiinteistönhoidon organisoinnin muutos vaikutti mm. ajokilometreihin, mutta erikoista oli se, että kiinteistöllä vietetty aikakin väheni. Liikeratojen käyminen Ulvilassa herättää myös mielenkiintoa. Onkohan puhelin ollut kotiin lähtiessä mukana ja sammutettu se vasta kotona. Toisaalta onko kotiin lähdetty liian ajoissa? Toimintaan aiheuttavia jatkoselvitys tarpeita tämä kuitenkin aiheuttaa. Visualisointi on erittäin selkeää ja voisi sanoa, että jopa tieteellistä taidetta.”

### **Visuaaliset menetelmät, Operation Dashboard -sovellus**

“Operation Dashboard on kaupungilla ollut käytössä lähinnä kokonaisuuden tiedottamiseen ja tarkkailuun. Uutena ideana tuli tuo kahden kartan käyttö samassa sovelluksessa ja tekemättömien tehtävänäntöjen irrallinen tarkkailu. Tämäkin otetaan käyttöön.”

### **Visuaaliset menetelmät, Liikeratojen 2D visualisointi**

“Tähän luokitteluun pitää osallistua työnjohtoa syystä, että toisilla on kovin paljon tekemistä, kun taas toisilla ei. Ristikkäiset liikeradat toisten kanssa ihmetyttää. Onkohan ne tuurausta toisissa kohteissa kuin normaalisti vai onko kohteiden resursoinnissa todella parannettavaa noin paljon. Liikeradoille pitäisi tehdä Operation Dashboard -sovellus, jossa voi luokitella yksittäisen työntekijän päivittäisen liikeradan minuutin tarkkuudella. Tästä olisi hyötyä muun muassa vakuutustapauksissa, kun selvitetään, onko esimerkiksi hiekoitusta hoidettu.”

### **Visuaaliset menetelmät, Liikeradat aika-avaruuskuutiassa**

“Tämä auttoi ymmärtämään, että on käytössä erilaisia käytäntöjä siitä, koska tehtävän tekeminen aloitetaan ja lopetetaan. Toisinaan kohteessa käydään vielä toisenkin kerran päivän aika. Pitää keskustella tekijöiden kanssa, miksi näin toimitaan. Analysointi auttoi ymmärtämään, että kiinteistöillä aikaa vietetään suhteellisen vähän aikaa. Pitää alkaa pohtimaan ratkaisua, miten kiinteistöillä vietettyä aikaa lisätään. Vanhanaikainen talonmieskäytäntö näkyy parhaiten sairaala kohteen visualisoinnissa. Aikaa siirtymisiin menee vähän, kun kaikki kohde rakennukset ovat samalla tontilla.”

### **Visuaaliset menetelmät, Aika-avaruuskuutioiden trendianalyysit**

“Vaikeaselkoinen analyysi ensimmäisissä kuvissa. Ehkä videotyyppinen voisi tuoda esiin asioita enemmän esimerkiksi viikko per animaatioidia. Sijaintipisteiden laskun syynä voi olla

eriäviä asioita, mutta toivoisin, että kun näitä analyysejä esitetään työntekijöille he ymmärtävät sijaintiseurannan merkityksen isossa mittakaavassa.”

### **Yleisesti, yhteenveto ja jatko**

“Palvelupyynnöjen(tehtävänannot) pitää saada osaksi rutiinitehtävien hoitoa. Rutiinit pitää jatkossa laittaa tehtävänannoksi. Näin saadaan rakastettua analyysiä ja mahdollisesti optimointia päivätöihin. Nyt tehtyjen analyysien jatkohyödyntäminen osaksi pitempiaikaista tutkimusta, jossa tutkitaan käytännön tehtävien tekemisen muutoksien vaikutusta reitteihin ja aikoihin. Analyysit mahdollistivat hyödyntämisen, kun on havaittu poikkeamia. Nyt muutos on mahdollista!”